

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.
OM protein - protein search, using sw model
Run on: February 11, 2006, 18:56:56 ; Search time 146.817 Seconds
(without alignments)
4414.224 Million cell updates/sec
Title: US-10-797-821-34
Perfect score: 7741
Sequence: 1 MDKKVRYKLRKVKGRVTVS.....DRYGRISYDANSGERVRIN 1475
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 2443163 seqs, 439378781 residues
Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : A_Geneseq_21.*
1: geneseqp1980s.*
2: geneseqp1990s.*
3: geneseqp2000s.*
4: geneseqp2001s.*
5: geneseqp2002s.*
6: geneseqp2003as.*
7: geneseqp2003bs.*
8: geneseqp2004s.*
9: geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7741	100.0	1475	5	Aau98027 S. mutans
2	7741	100.0	1475	7	Add93654 Streptoco
3	7741	100.0	1475	9	Adx37277 Streptoco
4	7740	99.9	1475	5	Aau98030 S. mutans
5	7737	99.9	1475	5	Aau98040 S. mutans
6	7736	99.9	1475	5	Aau98031 S. mutans
7	7735	99.9	1475	5	Aau98033 S. mutans
8	7734	99.9	1475	5	Aau98032 S. mutans
9	7729	99.8	1475	5	Aau98035 S. mutans
10	7729	99.8	1475	5	Aau98034 S. mutans
11	7727	99.8	1475	5	Aau98036 S. mutans
12	7721	99.7	1475	5	Aau98037 S. mutans
13	7714	99.7	1475	5	Aau98039 S. mutans
14	7711	99.6	1475	5	Aau98038 S. mutans
15	7628.5	98.5	1476	5	Aau79284 Streptoco
16	5285	68.3	1375	5	Aau98028 S. mutans
17	5285	68.3	1375	5	Aau79288 Streptoco
18	5285	68.3	1375	5	Add93655 Streptoco
19	5285	68.3	1375	9	Adx37278 Streptoco
20	5154.5	66.6	1017	5	Aau79285 Streptoco
21	4580	59.2	1590	7	Add93657 Streptoco
22	4580	59.2	1590	9	Adx37280 Streptoco
23	4531	58.5	1592	2	Aar32925 Glucosylt
24	3829.5	49.5	1430	5	Aau98044 S. mutans

ALIGNMENTS

RESULT 1
AAU98027
ID AAU98027 standard; protein; 1475 AA.
XX AC AAU98027;
XX DT 27-AUG-2002 (first entry)
XX DE S. mutans glucosyltransferase GTFB.
XX KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture.
XX OS Streptococcus mutans.
XX PN US2002031826-A1.
XX PD 14-MAR-2002.
XX PF 19-DEC-2000; 2000US-00740274.
XX PR 07-JUN-1995; 95US-00478704.
PR 07-JUN-1995; 95US-00482711.
PR 07-JUN-1995; 95US-00485243.
PR 16-JAN-1998; 98US-00007999.
PR 16-JAN-1998; 98US-00008172.
PR 20-JAN-1998; 98US-00009620.
PR 11-DEC-1998; 98US-00210361.
(NICH//) NICHOLS S E.
XX PI Nichols SE;
XX WPI; 2002-414332/44.
XX N-PSDB; ABK52938.
PT Glucosyltransferase B or D protein useful for producing a glucan useful
PT as substitutes for and additions to modified starch and latexes in paper
XX manufacture, comprises mutations in specific positions.
XX PS Disclosure; Page 21-25; 4pp; English.
XX CC The invention an isolated protein comprising a glucosyltransferase (GTF)
CC B polypeptide having changes at position from 1448V, D457N, D567T,
CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/K1014T,
CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a

CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
CC complementary polynucleotide, a ribonucleic acid sequence encoding the
CC GTF mutant, an expression cassette comprising the polynucleotide operably
CC linked to a promoter, a vector comprising the expression cassette, host
CC cell introduced with the vector, a transgenic plant comprising the
CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
CC coating composition comprising a glucan produced in a plant transformed
CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
CC thermoplastic molecule or their combinations or glucan and starch where
CC the glucan is produced in the amyloplast and/or vacuole or a maize line
CC deficient in starch biosynthesis, transformed with a gene encoding a
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
CC comprising the glucan (paper sizing/coating agent). The vector is useful
CC for producing a glucan in a plant. The method comprises transforming a
CC plant cell with the vector, growing the plant cell under plant growing
CC conditions to produce a regenerated plant and inducing expression of the
CC polynucleotide for a time sufficient to produce the glucan in the
CC regenerated plant, where the vector contains a transit sequence from
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
CC is produced in the amyloplast of potato or the vacuole of sugar beet.
CC Glucans are useful as substitutes for and additions to modified starch
CC and latexes in paper manufacture. Unlike prior art techniques, which
CC require input materials that produce chemical effluents, paper
CC manufacture utilising the glucan produced by GTF, which utilises
CC biologically produced input materials, is more cost-effective and
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
CC properties and impart gloss to the paper during coating step. The present
CC sequence represents GTFB
XX
SQ Sequence 1475 AA;

Query Match 100.0%; Score 7741; DB 5; Length 1475;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1475; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MDKKRYKLRKVKRWTVSVASAVMTLTLSGGLVKADSNESKQISNDNSTSVVTANE 60
DB 1 MDKKRYKLRKVKRWTVSVASAVMTLTLSGGLVKADSNESKQISNDNSTSVVTANE 60
QY 61 ESNVTEATSKQEAASSQTNHTVTSSSTSVVNPKEVVSPPYVGTASNGEKLQNTT 120
DB 61 ESNVTEATSKQEAASSQTNHTVTSSSTSVVNPKEVVSPPYVGTASNGEKLQNTT 120
QY 121 TVDKTSEAAANNISQTTTEADTDVDDSNAAANLQILEKLPNVKEIDGKYIYDNNKQVRT 180
DB 121 TVDKTSEAAANNISQTTTEADTDVDDSNAAANLQILEKLPNVKEIDGKYIYDNNKQVRT 180
QY 181 NFTLIADGKILHFDGTGAYTDSIDTVNKDITVTRSNLYKKYNQVYDRSAQSFEHVDHYL 240
DB 181 NFTLIADGKILHFDGTGAYTDSIDTVNKDITVTRSNLYKKYNQVYDRSAQSFEHVDHYL 240
QY 241 TAESWYRPKYILKOGKWTQSTOKDFRPLMTWPDQETORQYVNMNAQLGINKTYDDT 300
DB 241 TAESWYRPKYILKOGKWTQSTOKDFRPLMTWPDQETORQYVNMNAQLGINKTYDDT 300
QY 301 SNQLQNLIAAATIQAKIEAKITLKNLTDWLRQTI SAFVKTSAMNSDSEKPPDDHLONGA 360
DB 301 SNQLQNLIAAATIQAKIEAKITLKNLTDWLRQTI SAFVKTSAMNSDSEKPPDDHLONGA 360
QY 361 VLYDNEGKLTYPANSNYRILNPTNQTGKDPRTYADNTTGGYEFLLANDVDNSNPVVQ 420
DB 361 VLYDNEGKLTYPANSNYRILNPTNQTGKDPRTYADNTTGGYEFLLANDVDNSNPVVQ 420
QY 421 AEQLNLWHLFNMFGNIYANDPDANFDSIRVDADVNDVADLLQIAGDYLKAAGIHKNDKA 480
DB 421 AEQLNLWHLFNMFGNIYANDPDANFDSIRVDADVNDVADLLQIAGDYLKAAGIHKNDKA 480
QY 481 ANDHLSILEAWSNDNTPYLIHDDGDNMINMDNKLRLSLFLSLAKPLNQRSGMNPITNSLV 540
DB 481 ANDHLSILEAWSNDNTPYLIHDDGDNMINMDNKLRLSLFLSLAKPLNQRSGMNPITNSLV 540

RESULT 2
ADD93654
ID ADD93654 standard; protein; 1475 AA.
XX
AC ADD93654;
XX
DT 29-JAN-2004 (first entry)

QY 541 NRTDDNAETAAPVPSYFIRAHDSVQDIIADIIKAEINPNVVGYSFTMEEIKKAEIYNK 600
DB 541 NRTDDNAETAAPVPSYFIRAHDSVQDIIADIIKAEINPNVVGYSFTMEEIKKAEIYNK 600
QY 601 DLLATEKKYTHYNTALSVALLLTNKSSVPRVYVYGDMDTDDGQYMAHKNTINYEAIETLKA 660
DB 601 DLLATEKKYTHYNTALSVALLLTNKSSVPRVYVYGDMDTDDGQYMAHKNTINYEAIETLKA 660
QY 661 RIKYVSGQAMRNQOVGNSEIITSVRYGKGALKATDTGDRTRTTSVAVIEGNNPSLRUK 720
DB 661 RIKYVSGQAMRNQOVGNSEIITSVRYGKGALKATDTGDRTRTTSVAVIEGNNPSLRUK 720
QY 721 ASDRVVNMGAHKNQAVRPLLLTTDNGIKAYHSDQEAAGLVRYTNDRGELIPTAADIKG 780
DB 721 ASDRVVNMGAHKNQAVRPLLLTTDNGIKAYHSDQEAAGLVRYTNDRGELIPTAADIKG 780
QY 781 YANPOVSGYLGWVPVGAALIKMFALRLARPHQOMASVHQNAALDSRVMFEGFSNFQAPA 840
DB 781 YANPOVSGYLGWVPVGAALIKMFALRLARPHQOMASVHQNAALDSRVMFEGFSNFQAPA 840
QY 841 TKKBEYTNVVIKXNDVDFEMAPQVVSSTGSLDSVITQNGYAFTRDVLGITS 900
DB 841 TKKBEYTNVVIKXNDVDFEMAPQVVSSTGSLDSVITQNGYAFTRDVLGITS 900
QY 901 KPNKYGTADDLVKAALKALHSKGIKVMADWVPDQMYAFPEKEVVTATRVDKYCTPVAGSQI 960
DB 901 KPNKYGTADDLVKAALKALHSKGIKVMADWVPDQMYAFPEKEVVTATRVDKYCTPVAGSQI 960
QY 961 KNTLYVVDGKSSGKQQAQKYGGAFLEELQAKYPELFARKQISTGVPMPPSVKIKOWSAKY 1020
DB 961 KNTLYVVDGKSSGKQQAQKYGGAFLEELQAKYPELFARKQISTGVPMPPSVKIKOWSAKY 1020
QY 1021 FNGNTILGRGAGYVLKQATNTYFNISDNKEINFLPKTLNQLDRAILKNEGTAY 1140
DB 1021 FNGNTILGRGAGYVLKQATNTYFNISDNKEINFLPKTLNQLDRAILKNEGTAY 1140
QY 1141 YGNDGRRYENGYQFMGVRHFNNGEMSVGLTVIDGQVQYFDEMGYQAKGKFTVTADGK 1200
DB 1141 YGNDGRRYENGYQFMGVRHFNNGEMSVGLTVIDGQVQYFDEMGYQAKGKFTVTADGK 1200
QY 1201 IRYFDKQSGNNMYRNFIEBEGKWLILGEDGAAVTSQTINGQHLYFRANGVQVKGFEVY 1260
DB 1201 IRYFDKQSGNNMYRNFIEBEGKWLILGEDGAAVTSQTINGQHLYFRANGVQVKGFEVY 1260
QY 1261 DHGRIYSYDGNSGDQIRNRFVRNAQOMQYFDDNNGYAVTGARTINGQHLVFRANGVQVK 1320
DB 1261 DHGRIYSYDGNSGDQIRNRFVRNAQOMQYFDDNNGYAVTGARTINGQHLVFRANGVQVK 1320
QY 1321 GEFTVTDYGRISYDGNSGDQIRNRFVRNAQOMQYFDDNNGYAVTGARTINGQHLVFRAN 1380
DB 1321 GEFTVTDYGRISYDGNSGDQIRNRFVRNAQOMQYFDDNNGYAVTGARTINGQHLVFRAN 1380
QY 1391 GVQVKGFEVTDHGRISYDGNSGDQIRNRFVRNAQOMQYFDDNNGYAVTGARTINGQHL 1440
DB 1391 GVQVKGFEVTDHGRISYDGNSGDQIRNRFVRNAQOMQYFDDNNGYAVTGARTINGQHL 1440
QY 1441 YFRANGVQVKGFEVTDYGRISYDGNSEGRVIRIN 1475
DB 1441 YFRANGVQVKGFEVTDYGRISYDGNSEGRVIRIN 1475

RESULT 3

ADX37277
ID ADX37277 standard; protein; 1475 AA.XX AC ADX37277;
XX

DT 21-APR-2005 (first entry)

XX Streptococcus mutant glucan binding protein B variant #6.
XXKW immunogenicity; immune stimulation; glucan binding protein-B;
KW microparticle; major histocompatibility complex; tooth disease.
XX

OS Streptococcus mutans.

XX US2005031633-A1.
XX

PD 10-FEB-2005.

XX 09-MAR-2004; 2004US-00797821.
XXPR 13-APR-1998; 98US-0081550P.
PRPR 08-JAN-1999; 99US-0115142P.
PRPR 12-APR-1999; 99US-00290049.
PRPR 07-MAR-2002; 2002US-0363209P.
PRPR 08-AUG-2002; 2002US-0402483P.
PRPR 07-MAR-2003; 2003US-00383930.
XXPA (SMIT/) SMITH D J.
PAXX (TAUB/) TAUBMAN M A.
XXPI Smith DJ, Taubman WA;
XXXX WPI; 2005-151644/16.
XXXX New composition comprising a fragment of a glucan binding protein-B
PT (GbpB) that binds to MHC class II protein, and a biocompatible
PT microparticle, useful for producing an antibody (claimed) for immunizing
PT mammals against dental caries.
XX

PS Claim 7; SEQ ID NO 34; 73pp; English.

XX The invention relates to a composition comprising a fragment of a glucan
CC binding protein-B (GbpB) and a biocompatible microparticle, where the
CC fragment binds to a major histocompatibility complex (MHC) class II
CC protein. The composition is useful for producing an antibody for
CC immunizing mammals against dental caries. This sequence corresponds to a
CC Streptococcus mutans GbpB protein of the invention.
XX

SQ Sequence 1475 AA;

Query Match 100.0%; Score 7741; DB 9; Length 1475;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1475; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDKKVRYKLRKYKRWVTVSVASAVMTLTTLTSLGGLVKADSNESKSQISNDSNTSVVTANE 60

DB 1 MDKKVRYKLRKYKRWVTVSVASAVMTLTTLTSLGGLVKADSNESKSQISNDSNTSVVTANE 60

QY 61 ESNVITEATSKQEAASSQTNHTVTSSSTSVVNPKEVSNPNYTVGETASNGEKLQNQTT 120

DB 61 ESNVITEATSKQEAASSQTNHTVTSSSTSVVNPKEVSNPNYTVGETASNGEKLQNQTT 120

QY 121 TVDKTSEAAANNISKQTTTADTDVDDSNAAQLILEKLPNVKEIDGKYVYDNNKGVRT 180

DB 121 TVDKTSEAAANNISKQTTTADTDVDDSNAAQLILEKLPNVKEIDGKYVYDNNKGVRT 180

QY 181 NPTLIADGKILHFEDEGTADTSDTVNKNKDITVTRSNLYKKYNQVYDRSAQSFHVHDYL 240

DB 181 NPTLIADGKILHFEDEGTADTSDTVNKNKDITVTRSNLYKKYNQVYDRSAQSFHVHDYL 240

QY 241 TAESWYRPKYILKDGKTWTQSTEKDFRPLLTWPDQETQRYVYVYMAQLGINKTYDDT 300
DB |||||
241 TAESWYRPKYILKDGKTWTQSTEKDFRPLLTWPDQETQRYVYVYMAQLGINKTYDDT 300
QY 301 SNQLQLNIAAATIQAIEBAKITTLKNTDMLRQTISAFVKTQSAMNSDSEKPPDDHLQNGA 360
DB |||||
301 SNQLQLNIAAATIQAIEBAKITTLKNTDMLRQTISAFVKTQSAMNSDSEKPPDDHLQNGA 360
QY 361 VLYDNEGKLTYPYANSNYRIILNRTPTNQTGKDPRTYADNTTGGYEFLLANDVDSNPVVQ 420
DB |||||
361 VLYDNEGKLTYPYANSNYRIILNRTPTNQTGKDPRTYADNTTGGYEFLLANDVDSNPVVQ 420
QY 421 AEQLNLHLFLMNFNGNIYANDPDANFDSIRVDVNDVADLLQIAGDYLKAAAGIHKNDKA 480
DB |||||
421 AEQLNLHLFLMNFNGNIYANDPDANFDSIRVDVNDVADLLQIAGDYLKAAAGIHKNDKA 480
QY 481 ANDHLSILEAWSNDPTPYLHDDGDNMINMDNKLRLSLFLSLAKPLNQRSGMNPILTNSLV 540
DB |||||
481 ANDHLSILEAWSNDPTPYLHDDGDNMINMDNKLRLSLFLSLAKPLNQRSGMNPILTNSLV 540
QY 541 NRTDNDASTAAVPSYSPFIRAHDSVODLIADIIKAEINPNVVGYSFTWEEIKKAEIYNK 600
DB |||||
541 NRTDNDASTAAVPSYSPFIRAHDSVODLIADIIKAEINPNVVGYSFTWEEIKKAEIYNK 600
QY 601 DLLATEKKYTHYNTALSVALLLTNKSSVPRVYVYGMFTDDGQYMAHKTINYEAIETLLKA 660
DB |||||
601 DLLATEKKYTHYNTALSVALLLTNKSSVPRVYVYGMFTDDGQYMAHKTINYEAIETLLKA 660
QY 661 RIKYVSGGOAMRNQOVGNSEIITSVRYGKGALKATDTGDRTRTTSVAVIEGNNPFLRLK 720
DB |||||
661 RIKYVSGGOAMRNQOVGNSEIITSVRYGKGALKATDTGDRTRTTSVAVIEGNNPFLRLK 720
QY 721 ASDRVVNMGAHKNQAYRPLLLTTDNGIKAYHSQBAAGLVRYTNDRGELIFTAADIKG 780
DB |||||
721 ASDRVVNMGAHKNQAYRPLLLTTDNGIKAYHSQBAAGLVRYTNDRGELIFTAADIKG 780
QY 781 YANPOVSGVLGVVVPVGAALIKMPALRLARPHQOQWASVHQAALDSRVNMFEGSFOAPA 840
DB |||||
781 YANPOVSGVLGVVVPVGAALIKMPALRLARPHQOQWASVHQAALDSRVNMFEGSFOAPA 840
QY 841 TKKEEYTNVIAKNVDKFAEWGVTDFEMAPQVSVSTSGSFLDSVIFQNGYAFPTDRYDLGIS 900
DB |||||
841 TKKEEYTNVIAKNVDKFAEWGVTDFEMAPQVSVSTSGSFLDSVIFQNGYAFPTDRYDLGIS 900
QY 901 KPNKGTADDLVAIKALHSGIKIWMADVPDQMTAFPEKEVVTATRDVKYGTPTVAGSOI 960
DB |||||
901 KPNKGTADDLVAIKALHSGIKIWMADVPDQMTAFPEKEVVTATRDVKYGTPTVAGSOI 960
QY 961 KNTLVVVDGKSSGKQQAQYKGAFLLEELQAKYPELFARKQISTGVPMDPSPVKIKOWSAKY 1020
DB |||||
961 KNTLVVVDGKSSGKQQAQYKGAFLLEELQAKYPELFARKQISTGVPMDPSPVKIKOWSAKY 1020
QY 1021 FNGTNILGRGAGYVLKDOATNTYFNI SDNKEINFLPKTLNLDQSDQVGSYDGKGVYVYST 1080
DB |||||
1021 FNGTNILGRGAGYVLKDOATNTYFNI SDNKEINFLPKTLNLDQSDQVGSYDGKGVYVYST 1080
QY 1081 SGYQAKNTFISEGDKWYIFDNNGVMTGAQISNGVNGYVYFSLNGLQIRDAILKNEDGTAY 1140
DB |||||
1081 SGYQAKNTFISEGDKWYIFDNNGVMTGAQISNGVNGYVYFSLNGLQIRDAILKNEDGTAY 1140
QY 1141 YGNDGRRYENGYYQFMSGVRHFNNGEMSVGLTVIDGVOYVDEMGYQAKGKFTVTADCK 1200
DB |||||
1141 YGNDGRRYENGYYQFMSGVRHFNNGEMSVGLTVIDGVOYVDEMGYQAKGKFTVTADCK 1200
QY 1201 IRYFDKQSGNMYRNFRIENEKGWLYLGEDGAATVTSQTINGQHLYFRANGVQVKGFEVY 1260
DB |||||
1201 IRYFDKQSGNMYRNFRIENEKGWLYLGEDGAATVTSQTINGQHLYFRANGVQVKGFEVY 1260
QY 1261 DHHGRISYDNGSGDQIRNRFVRNAQOQWFFDNNGYAVTGARTINGQLLYFRANGVQVK 1320
DB |||||
1261 DHHGRISYDNGSGDQIRNRFVRNAQOQWFFDNNGYAVTGARTINGQLLYFRANGVQVK 1320
QY 1321 GEFVTDTRYGRISYDNGSGDQIRNRFVRNAQOQWFFDNNGYAVTGARTINGQHLYFRAN 1380

Db 1321 GSFVTDYGRISYDNGSDQIRNRFVRNAQGWYFDNNGYAVTGARTINGQHLYFRAN 1380
Qy 1381 GQVQKGEFVTDHGRISYDNGSDQIRNRFVRNAQGWYFDNNGYAVTGARTINGQHL 1440
Db 1381 GQVQKGEFVTDHGRISYDNGSDQIRNRFVRNAQGWYFDNNGYAVTGARTINGQHL 1440
Qy 1441 YFRANQVQKGFVTDYGRISYDNGSERVRAIN 1475
Db 1441 YFRANQVQKGFVTDYGRISYDNGSERVRAIN 1475

RESULT 4
AAU98030
ID AAU98030 standard; protein; 1475 AA.
XX
AC AAU98030;
XX
DT 27-AUG-2002 (first entry)
XX
S. mutans glucosyltransferase GTFB mutant I448V.
XX
DE Glucosyltransferase; GTFB; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture; mutant; mutein.
XX
OS Streptococcus mutans.
OS Synthetic.

PH Key Location/Qualifiers
FT Misc-difference 448 /note= "Wild-type Ile substituted by Val"
FT XX
PN US2002031826-A1.
XX
PD 14-MAR-2002.
XX
PF 19-DEC-2000; 2000US-00740274.
XX
PR 07-JUN-1995; 95US-00478704.
PR 07-JUN-1995; 95US-00482711.
PR 07-JUN-1995; 95US-00485243.
PR 16-JAN-1998; 98US-00007999.
PR 16-JAN-1998; 98US-00008172.
PR 20-JAN-1998; 98US-00009620.
PR 11-DEC-1998; 98US-00210361.
XX
PA (NICH/) NICHOLS S E.
XX
PI Nichols SE;
XX
WPI; 2002-414332/44.

Glucosyltransferase B or D protein useful for producing a glucan useful as substitutes for and additions to modified starch and latexes in paper manufacture, comprises mutations in specific positions.
XX
PS Claim 36; Page; 44pp; English.

XX The invention an isolated protein comprising a glucosyltransferase (GTF) B polypeptide having changes at position from I448V, D457N, D567T, K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T, I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a GTF D polypeptide having changes at positions from T589D, T589E, N471D, N471D/T589D, and N471D/T589E. Also included are a glucan produced by the GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its complementary polynucleotide, a ribonucleic acid sequence encoding the GTF mutant, an expression cassette comprising the polynucleotide operably linked to a promoter, a vector comprising the expression cassette, host cell introduced with the vector, a transgenic plant comprising the vector, a seed or tuber from the transgenic plant, a paper sizing and/or coating composition comprising a glucan produced in a plant transformed with a gene encoding the mutant GTF, wild type or, starch, a latex,

CC thermoplastic molecule or their combinations or glucan and starch where the glucan is produced in the amyloplast and/or vacuole or a maize line deficient in starch biosynthesis, transformed with a gene encoding a glucosyltransferase B or D enzyme, wild-type or mutant and a paper comprising the glucan (paper sizing/coating agent). The vector is useful for producing a glucan in a plant. The method comprises transforming a plant cell with the vector, growing the plant cell under plant growing conditions to produce a regenerated plant and inducing expression of the polynucleotide for a time sufficient to produce the glucan in the regenerated plant, where the vector contains a transit sequence from ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and chlorophyll AB binding protein to produce a transgenic plant, and glucan is produced in the amyloplast of potato or the vacuole of sugar beet. Glucans are useful as substitutes for and additions to modified starch and latexes in paper manufacture. Unlike prior art techniques, which require input materials that produce chemical effluents, paper manufacture utilising the glucan produced by GTF, which utilises biologically produced input materials, is more cost-effective and environmentally friendly. Moreover, glucans also exhibit thermoplastic properties and impart gloss to the paper during coating step. The present sequence represents a GTFB mutant of the invention. Note: The present sequence is not shown in the specification but was created by the indexer CC using the GTFB sequence appearing as AAU98027 and the information in CC claim 36
XX

SQ Sequence 1475 AA;

Query Match 99.9%; Score 7740; DB 5; Length 1475;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1474; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MDKVRVYKLRVKYKRWVTVSVASAVMTLTTLGGVLKADSNESKQISNDSNTSVVTANE 60
Db 1 MDKVRVYKLRVKYKRWVTVSVASAVMTLTTLGGVLKADSNESKQISNDSNTSVVTANE 60
Qy 61 ESNVITEATSKQEAASSQTNHTVTSSSTSVVNPKEVSVNPTVGTETASNGEKLQNTT 120
Db 61 ESNVITEATSKQEAASSQTNHTVTSSSTSVVNPKEVSVNPTVGTETASNGEKLQNTT 120
Qy 121 TVDKTSEAAANNISKQTTADTDVDDSNAAANLQILEKLPNVKEIDGKYVYVNNNGKVRT 180
Db 121 TVDKTSEAAANNISKQTTADTDVDDSNAAANLQILEKLPNVKEIDGKYVYVNNNGKVRT 180
Qy 181 NFTLIADGKILHFDGTGAYTDTSIDTVNKDI VTTTRSNLYKKYNQVYDRAQSFHVDHYL 240
Db 181 NFTLIADGKILHFDGTGAYTDTSIDTVNKDI VTTTRSNLYKKYNQVYDRAQSFHVDHYL 240
Qy 241 TAEVYRPRKYLKQGTWVTSQTEKDFRPLMTWPDQETQRYVNMNAQLGINKTYDDT 300
Db 241 TAEVYRPRKYLKQGTWVTSQTEKDFRPLMTWPDQETQRYVNMNAQLGINKTYDDT 300
Qy 301 SNQLQNTAAATIAQIEAKITTLKNTDMLROTISAFVKTQSAWNSDSEKPPDDHLONGA 360
Db 301 SNQLQNTAAATIAQIEAKITTLKNTDMLROTISAFVKTQSAWNSDSEKPPDDHLONGA 360
Qy 361 VLYDNEGKLTYPYANSNYRILNRTPTNQTKGKDPRTADNTIGGYEFLANDVDSNPVYQ 420
Db 361 VLYDNEGKLTYPYANSNYRILNRTPTNQTKGKDPRTADNTIGGYEFLANDVDSNPVYQ 420
Qy 421 AEQLNLWHLFMNFGNIYANDPDANFDSIRVDAVDNVDADLLQIAGDYKAAKGIIHKNDKA 480
Db 421 AEQLNLWHLFMNFGNIYANDPDANFDSIRVDAVDNVDADLLQIAGDYKAAKGIIHKNDKA 480
Qy 481 ANDHLSIILEANSNDNTPYLHDDGDNMINMDNKLRLSLFLSLAKPLNQRSGMPLTNSLV 540
Db 481 ANDHLSIILEANSNDNTPYLHDDGDNMINMDNKLRLSLFLSLAKPLNQRSGMPLTNSLV 540
Qy 541 NRTDDNAETAAVPSYFTRAHDSEVDQIADI IKAENPNVVGYSFTWEEI KKAPEIYNK 600
Db 541 NRTDDNAETAAVPSYFTRAHDSEVDQIADI IKAENPNVVGYSFTWEEI KKAPEIYNK 600
Qy 601 DLLATEKKTHTYNALSVALLLTNKSVPRVYVYVGMFTDDGQYMAHKTINYEAIETLLKA 660
Db 601 DLLATEKKTHTYNALSVALLLTNKSVPRVYVYVGMFTDDGQYMAHKTINYEAIETLLKA 660

Db 601 DLLATEKKYTHYNTALSYALLTNKSSVPRVYVYGMFTDDGQYMAHKNTINYEAIETLLKA 660
Qy 661 RIKYVSGGAMRNOQVGNSEIITSVRYGKALKATDGTDRTRTSGVAVIEGNPSLRK 720
Db 661 RIKYVSGGAMRNOQVGNSEIITSVRYGKALKATDGTDRTRTSGVAVIEGNPSLRK 720
Qy 721 ASDRVVNVNGAAHKNQARPLLLTTDNGIKAVHSDQEAAGLVRYTNDRGELIFTAADIKG 780
Db 721 ASDRVVNVNGAAHKNQARPLLLTTDNGIKAVHSDQEAAGLVRYTNDRGELIFTAADIKG 780
Qy 781 YANPOVSGYLVGVVPGAALIKWFAIRLARPHQOMASVHQNAALDSRVMEFSGSFQAF 840
Db 781 YANPOVSGYLVGVVPGAALIKWFAIRLARPHQOMASVHQNAALDSRVMEFSGSFQAF 840
Qy 841 TKKEBYTNVIAKNVDKFAENGVTDFEMAPQVVSSTGDSFLDSVTQNGYAFTRDRYDLGIS 900
Db 841 TKKEBYTNVIAKNVDKFAENGVTDFEMAPQVVSSTGDSFLDSVTQNGYAFTRDRYDLGIS 900
Qy 901 KPNKYGTADDLVKAIKALHSKGIKYMADWVPDMYAFPEKEVVTATRVVDKYGTPVAGSQI 960
Db 901 KPNKYGTADDLVKAIKALHSKGIKYMADWVPDMYAFPEKEVVTATRVVDKYGTPVAGSQI 960
Qy 961 KNTLVVVDGSKGDKQQAQYKGAFLLEELQAKYPFELFARKQISTGVPMPSVKIKQWSAKY 1020
Db 961 KNTLVVVDGSKGDKQQAQYKGAFLLEELQAKYPFELFARKQISTGVPMPSVKIKQWSAKY 1020
Qy 1021 FNGTNILGRGAGYVLKQATNTYFNI SDNKEINFLPKTLNODSQVFSYDGKGVVYST 1080
Db 1021 FNGTNILGRGAGYVLKQATNTYFNI SDNKEINFLPKTLNODSQVFSYDGKGVVYST 1080
Qy 1081 SGYQAKNTFISGDKWYFDNNGYVMTGAQSNNGVYFLSNGQLRLDAILKNEGDGVAY 1140
Db 1081 SGYQAKNTFISGDKWYFDNNGYVMTGAQSNNGVYFLSNGQLRLDAILKNEGDGVAY 1140
Qy 1141 YGNDGRRYENGYYQFMSGWRHFNNGEMSVGLTVIDGVOYFDEMGYQAKGFVTTADGK 1200
Db 1141 YGNDGRRYENGYYQFMSGWRHFNNGEMSVGLTVIDGVOYFDEMGYQAKGFVTTADGK 1200
Qy 1201 TRYFDKQSGNMYRNFRIENEGKWLVLGEGDGAATVTSOTINGHLYFRANGVQVKGFPVT 1260
Db 1201 TRYFDKQSGNMYRNFRIENEGKWLVLGEGDGAATVTSOTINGHLYFRANGVQVKGFPVT 1260
Qy 1261 DHGGRISYDNGSGDQIRNRFVRNAQGWFFYFDNNGYAVTGARTINGOLLYFRANGVQVK 1320
Db 1261 DHGGRISYDNGSGDQIRNRFVRNAQGWFFYFDNNGYAVTGARTINGOLLYFRANGVQVK 1320
Qy 1321 GEFVTDYGRISYDNGSGDQIRNRFVRNAQGWFFYFDNNGYAVTGARTINGOHLYFRAN 1380
Db 1321 GEFVTDYGRISYDNGSGDQIRNRFVRNAQGWFFYFDNNGYAVTGARTINGOHLYFRAN 1380
Qy 1381 GVQVKGFPVTDYGRISYDNGSGDQIRNRFVRNAQGWFFYFDNNGYAVTGARTINGOHL 1440
Db 1381 GVQVKGFPVTDYGRISYDNGSGDQIRNRFVRNAQGWFFYFDNNGYAVTGARTINGOHL 1440
Qy 1441 YFRANGVQVKGFPVTDYGRISYDNGSERVRIN 1475
Db 1441 YFRANGVQVKGFPVTDYGRISYDNGSERVRIN 1475

RESULT 5

AAU98040

ID AAU98040 standard; protein; 1475 AA.

XX AC AAU98040;

XX DT 27-AUG-2002 (first entry)

XX DE S. mutans glucosyltransferase GTFB mutant K779Q.

XX KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;

XX KW coating composition; glucan; starch; latex; thermoplastic molecule;

XX KW amyloplast; vacuole; paper manufacture; mutant; mutein.

OS Streptococcus mutans.
OS Synthetic.

Key Location/Qualifiers

FT Misc-difference 779 /note= "Wild-type Lys substituted by Gln"

XX US2002031826-A1.

XX 14-MAR-2002.

XX 19-DEC-2000; 2000US-00740274.

XX 07-JUN-1995; 95US-00478704.

XX 07-JUN-1995; 95US-00482711.

XX 16-JAN-1998; 98US-00007999.

XX 16-JAN-1998; 98US-00008172.

XX 20-JAN-1998; 98US-00009620.

XX 11-DEC-1998; 98US-00210361.

XX (NICH/) NICHOLS S E.

XX Nichols SE;

XX WPI; 2002-414332/44.

XX Glucosyltransferase B or D protein useful for producing a glucan useful

as substitutes for and additions to modified starch and latexes in paper

manufacture, comprises mutations in specific positions.

XX Claim 36; Page; 44pp; English.

XX The invention an isolated protein comprising a glucosyltransferase (GTF)

B polypeptide having changes at position from 1448V, D457N, D567T,

K1014T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,

I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a

GTF D polypeptide having changes at positions from T589D, T589E, N471D,

N471D/T589D, and N471D/T589E. Also included are a glucan produced by the

GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its

complementary polynucleotide, a ribonucleic acid sequence encoding the

GTF mutant, an expression cassette comprising the polynucleotide operably

linked to a promoter, a vector comprising the expression cassette, host

cell introduced with the vector, a transgenic plant comprising the

vector, a seed or tuber from the transgenic plant, a paper sizing and/or

coating composition comprising a glucan produced in a plant transformed

with a gene encoding the mutant GTF, wild type or, starch, a latex,

thermoplastic molecule or their combinations or glucan and starch where

the glucan is produced in the amyloplast and/or vacuole or a maize line

deficient in starch biosynthesis, transformed with a gene encoding a

glucosyltransferase B or D enzyme, wild-type or mutant and a paper

comprising the glucan (paper sizing/coating agent). The vector is useful

for producing a glucan in a plant. The method comprises transforming a

plant cell with the vector, growing the plant cell under plant growing

conditions to produce a regenerated plant and inducing expression of the

polynucleotide for a time sufficient to produce the glucan in the

regenerated plant, where the vector contains a transit sequence from

ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and

chlorophyll AB binding protein to produce a transgenic plant, and glucan

is produced in the amyloplast of potato or the vacuole of sugar beet.

Glucans are useful as substitutes for and additions to modified starch

and latexes in paper manufacture. Unlike prior art techniques, which

require input materials that produce chemical effluents, paper

manufacture utilising the glucan produced by GTF, which utilises

biologically produced input materials, is more cost-effective and

environmentally friendly. Moreover, glucans also exhibit thermoplastic

properties and impart gloss to the paper during coating step. The present

sequence represents a GTFB mutant of the invention. Note: The present

sequence is not shown in the specification but was created by the indexer

using the GTFB sequence appearing as AAU98027 and the information in

claim 36

XX Sequence 1475 AA;

XX Glucosyltransferase B or D protein useful for producing a glucan useful
PT as substitutes for and additions to modified starch and latexes in paper
PT manufacture, comprises mutations in specific positions.
XX
PS
XX
XX

Claim 36; Page; 44pp; English.

CC The invention an isolated protein comprising a glucosyltransferase (GTF)
CC B polypeptide having changes at position from 1448V, D457N, D567T,
CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T,
CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
CC complementary polynucleotide, a ribonucleic acid sequence encoding the
CC GTF mutant, an expression cassette comprising the polynucleotide operably
CC linked to a promoter, a vector comprising the expression cassette, host
CC cell introduced with the vector, a transgenic plant comprising the
CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
CC coating composition comprising a glucan produced in a plant transformed
CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
CC thermoplastic molecule or their combinations or glucan and starch where
CC the glucan is produced in the amyloplast and/or vacuole or a maize line
CC deficient in starch biosynthesis, transformed with a gene encoding a
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
CC comprising the glucan (paper sizing/coating agent). The vector is useful
CC for producing a glucan in a plant. The method comprises transforming a
CC plant cell with the vector, growing the plant cell under plant growing
CC conditions to produce a regenerated plant and inducing expression of the
CC polynucleotide for a time sufficient to produce the glucan in the
CC regenerated plant, where the vector contains a transit sequence from
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
CC is produced in the amyloplast of potato or the vacuole of sugar beet.
CC Glucans are useful as substitutes for and additions to modified starch
CC and latexes in paper manufacture. Unlike prior art techniques, which
CC require input materials that produce chemical effluents, paper
CC manufacture utilising the glucan produced by GTF, which utilises
CC biologically produced input materials, is more cost-effective and
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
CC properties and impart gloss to the paper during coating step. The present
CC sequence represents a GTFB mutant of the invention. Note: The present
CC sequence is not shown in the specification but was created by the indexer
CC using the GTFB sequence appearing as AAU98027 and the information in
CC claim 36
XX

SQ Sequence 1475 AA;

Query Match 99.9%; Score 7736; DB 5; Length 1475;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1474; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MDKVRKYLKRVKKEWTVSVASAVMTLTLSGGLVKADSNESKSQISNDNNTSVVTANE 60
DB 1 MDKVRKYLKRVKKEWTVSVASAVMTLTLSGGLVKADSNESKSQISNDNNTSVVTANE 60
QY 61 ESNVITEATSKQEAASSQTNHTVTTSSSTSVVNPKEVSNPYTVGTASNGEKLQNTT 120
DB 61 ESNVITEATSKQEAASSQTNHTVTTSSSTSVVNPKEVSNPYTVGTASNGEKLQNTT 120
QY 121 TVDKTSEAAANNISKQTTTADTDVDDSNAAANLQILEKLPNVEIDGKYYYDDNNGKVRT 180
DB 121 TVDKTSEAAANNISKQTTTADTDVDDSNAAANLQILEKLPNVEIDGKYYYDDNNGKVRT 180
QY 181 NFTLIADGKILHFDSTGAYTDTSDTVNKDIIVTTRSNLYKKYNQVYDRSAQSFEHVDHYL 240
DB 181 NFTLIADGKILHFDSTGAYTDTSDTVNKDIIVTTRSNLYKKYNQVYDRSAQSFEHVDHYL 240
QY 241 TAESWYRPKYILKDGKTKTQSTTEKDFRPLMTWPDQETQRQYVNMNAQLGINKTYDDT 300
DB 241 TAESWYRPKYILKDGKTKTQSTTEKDFRPLMTWPDQETQRQYVNMNAQLGINKTYDDT 300
QY 301 SNQLQNTAAATIOAKIEAKITTLKNTDWRQTISAFVKTSQSAWNSDSEKPPDDHLQNGA 360

DB 301 SNQLQNTAAATIOAKIEAKITTLKNTDWRQTISAFVKTSQSAWNSDSEKPPDDHLQNGA 360
QY 361 VLYNDEGKLTTPYANSNYRILNRTPTNQTKXDPRTYADNTTIGGYEFLLANDVDSNPVVQ 420
DB 361 VLYNDEGKLTTPYANSNYRILNRTPTNQTKXDPRTYADNTTIGGYEFLLANDVDSNPVVQ 420
QY 421 AEOLNWLHFLMNFNFIYANDPDANFDSIRVDAVDNVDADLLQIAGDYLKAAKGIHNKDKA 480
DB 421 AEOLNWLHFLMNFNFIYANDPDANFDSIRVDAVDNVDADLLQIAGDYLKAAKGIHNKDKA 480
QY 481 ANDHLSILEAASDNDTPYLHDDGDNMINDNKLRLSLFLSLAKPLNQSGMNPILTNSLV 540
DB 481 ANDHLSILEAASDNDTPYLHDDGDNMINDNKLRLSLFLSLAKPLNQSGMNPILTNSLV 540
QY 541 NRTDDNAETAAPSVSFIKRAHSDVQDLIADIKAIEINPNVVGVSFTMEEIKKAFIYNK 600
DB 541 NRTDDNAETAAPSVSFIKRAHSDVQDLIADIKAIEINPNVVGVSFTMEEIKKAFIYNK 600
QY 601 DLLATEKYYTHNTALSVALLLTNKSSVPRVYVYGMFTDDGOYMAHKTINYEAIETLLKA 660
DB 601 DLLATEKYYTHNTALSVALLLTNKSSVPRVYVYGMFTDDGOYMAHKTINYEAIETLLKA 660
QY 661 RIKVYSGQAMRNQOVGNSEIITSVRYCKGKALKATDGTDRTRTTSVAVIEGNNSLRUK 720
DB 661 RIKVYSGQAMRNQOVGNSEIITSVRYCKGKALKATDGTDRTRTTSVAVIEGNNSLRUK 720
QY 721 ASDRVVNMGAHKNQAYRPLLLTTDNGIKAYHSDQEAAGLVRYTNRGELIFTAADIKG 780
DB 721 ASDRVVNMGAHKNQAYRPLLLTTDNGIKAYHSDQEAAGLVRYTNRGELIFTAADIKG 780
QY 781 YANPQVSGYLGVWVPVGAALIKMFALRLARPHQQWASHVQNAALDSRVNPEGSFQAPA 840
DB 781 YANPQVSGYLGVWVPVGAALIKMFALRLARPHQQWASHVQNAALDSRVNPEGSFQAPA 840
QY 841 TKKEEYTNVITAKNVDFEAGVTDFFENAPQVSVSTDSGLDSVITQNGVAFTRDYLGTG 900
DB 841 TKKEEYTNVITAKNVDFEAGVTDFFENAPQVSVSTDSGLDSVITQNGVAFTRDYLGTG 900
QY 901 KPNYGTADDLVKAIKALHSGIKVMADWPDQMYAFPEKEVVTATRVDKYGTVPVAGSQI 960
DB 901 KPNYGTADDLVKAIKALHSGIKVMADWPDQMYAFPEKEVVTATRVDKYGTVPVAGSQI 960
QY 961 KNTLYVVDGKSSGKQQAQYKGAFLLELOAKYPELFAKQISTGVPMDFSVKIKWSAKY 1020
DB 961 KNTLYVVDGKSSGKQQAQYKGAFLLELOAKYPELFAKQISTGVPMDFSVKIKWSAKY 1020
QY 1021 FNGTNILGRGAGYVLKQDQATNTYFNI SDNKEINFLPKTLLNQDSOVGFSYDGKGVYVYST 1080
DB 1021 FNGTNILGRGAGYVLKQDQATNTYFNI SDNKEINFLPKTLLNQDSOVGFSYDGKGVYVYST 1080
QY 1081 SGYQAKNTFISEGDKWYVFDNNGYVMTGAQISNGVYVYFLSNGQLRDAILKNEGTAY 1140
DB 1081 SGYQAKNTFISEGDKWYVFDNNGYVMTGAQISNGVYVYFLSNGQLRDAILKNEGTAY 1140
QY 1141 YGNDGRRYENGYYQPMGSGVWRHFNNGEMSVGLTVIDGQVQFDEMGYQAKGKFTVTDGK 1200
DB 1141 YGNDGRRYENGYYQPMGSGVWRHFNNGEMSVGLTVIDGQVQFDEMGYQAKGKFTVTDGK 1200
QY 1201 IRYFDKQSGNNYRNFRIENEBEGKWLILGEDGAAVTGSOTINGOHLYFRANGVQVGEFVT 1260
DB 1201 IRYFDKQSGNNYRNFRIENEBEGKWLILGEDGAAVTGSOTINGOHLYFRANGVQVGEFVT 1260
QY 1261 DHGRIISYDNGSGDQIRNRFVRNAQGWFFDNNGYAVTGARTINGQLLYFRANGVQVK 1320
DB 1261 DHGRIISYDNGSGDQIRNRFVRNAQGWFFDNNGYAVTGARTINGQLLYFRANGVQVK 1320
QY 1321 GEFVTDYGRISYDNGSGDQIRNRFVRNAQGWFFDNNGYAVTGARTINGOHLYFRAN 1380
DB 1321 GEFVTDYGRISYDNGSGDQIRNRFVRNAQGWFFDNNGYAVTGARTINGOHLYFRAN 1380
QY 1381 GVQVKGFEVTDYGRISYDNGSGDQIRNRFVRNAQGWFFDNNGYAVTGARTINGOHL 1440

Db 1381 GVOVKGEFVTDHGRISYYDGNSSGDOIRNFRVNRAGQWFYEDNNGYAVTGARTINGOHL 1440

Qy 1441 YFRANGVQVKGFEVTDYGRISYDANSGERVRIN 1475
 |||||

Db 1441 YFRANGVQVKGFEVTDYGRISYDANSGERVRIN 1475
 |||||

RESULT 7
 AAU98033
 ID AAU98033 standard; protein; 1475 AA.
 AC AAU98033;
 XX
 XX 27-AUG-2002 (first entry)
 XX
 DE S. mutans glucosyltransferase GTFB mutant K1014T.
 XX
 KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;
 KW coating composition; glucan; starch; latex; thermoplastic molecule;
 KW amyloplast; vacuole; paper manufacture; mutant; mutein.
 XX
 OS Streptococcus mutans.
 OS Synthetic.
 XX
 XX Key Location/Qualifiers
 FH Misc-difference 1014
 FT /note= "Wild-type Lys substituted by Thr"
 XX
 XX US2002031826-A1.
 XX
 PD 14-MAR-2002.
 XX
 XX 19-DEC-2000; 2000US-00740274.
 XX
 PR 07-JUN-1995; 95US-00478704.
 PR 07-JUN-1995; 95US-00482711.
 PR 07-JUN-1995; 95US-00485243.
 PR 16-JAN-1998; 98US-00007999.
 PR 16-JAN-1998; 98US-00008172.
 PR 20-JAN-1998; 98US-00009620.
 PR 11-DEC-1998; 98US-00210361.
 XX
 PA (NICH//) NICHOLS S E.
 XX
 PI Nichols SE;
 XX
 XX WPI; 2002-414382/44.
 DR
 XX
 PT Glucosyltransferase B or D protein useful for producing a glucan useful
 PT as substitutes for and additions to modified starch and latexes in paper
 PT manufacture, comprises mutations in specific positions.
 PT
 XX
 PS Claim 36; Page; 44pp; English.
 XX

The invention an isolated protein comprising a glucosyltransferase (GTF) B polypeptide having changes at position from 1448V, D457N, D567T, K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T, I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a GTF D polypeptide having changes at positions from T589D, T589E, N471D, N471D/T589D, and N471D/T589E. Also included are a glucan produced by the GTF mutant, an isolated polynucleotide which encodes p1 or p2, or its complementary polynucleotide, a ribonucleic acid sequence encoding the GTF mutant, an expression cassette comprising the polynucleotide operably linked to a promoter, a vector comprising the expression cassette, host cell introduced with the vector, a transgenic plant comprising the vector, a seed or tuber from the transgenic plant, a paper sizing and/or coating composition comprising a glucan produced in a plant transformed with a gene encoding the mutant GTF, wild type or, starch, a latex, thermoplastic molecule or their combinations or glucan and starch where the glucan is produced in the amyloplast and/or vacuole or a maize line deficient in starch biosynthesis, transformed with a gene encoding a glucosyltransferase B or D enzyme, wild-type or mutant and a paper comprising the glucan (paper sizing/coating agent). The vector is useful

CC for producing a glucan in a plant. The method comprises transforming a plant cell with the vector, growing the plant cell under plant growing conditions to produce a regenerated plant and inducing expression of the polynucleotide for a time sufficient to produce the glucan in the regenerated plant, where the vector contains a transit sequence from ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and chlorophyll AB binding protein to produce a transgenic plant, and glucan is produced in the amyloplast of potato or the vacuole of sugar beet. Glucans are useful as substitutes for and additions to modified starch and latexes in paper manufacture. Unlike prior art techniques, which require input materials that produce chemical effluents, paper manufacture utilising the glucan produced by GTF, which utilises biologically produced input materials, is more cost-effective and environmentally friendly. Moreover, glucans also exhibit thermoplastic properties and impart gloss to the paper during coating step. The present sequence represents a GTFB mutant of the invention. Note: The present sequence is not shown in the specification but was created by the indexer using the GTFB sequence appearing as AAU98027 and the information in claim 36

XX
 SQ Sequence 1475 AA;

Query Match 99.9%; Score 7735; DB 5; Length 1475;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1474; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MDKVRVYKLRVKRWVTVSVASAVMTLTTLGGLVKADSNESKQISNDSTSVVTANE 60
 |||||

Db 1 MDKVRVYKLRVKRWVTVSVASAVMTLTTLGGLVKADSNESKQISNDSTSVVTANE 60
 |||||

Qy 61 ESNVITEATSKQEAASSQTNHTVTTSSTSVVNPKEVVSNPYTVGTASNGEKLQNTT 120
 |||||

Db 61 ESNVITEATSKQEAASSQTNHTVTTSSTSVVNPKEVVSNPYTVGTASNGEKLQNTT 120
 |||||

Qy 121 TVDKTSEAAANNISKQTTTEADTDVDDSNAAANLQLEKLPNKVEIDGKYIYDNNKQVPT 180
 |||||

Db 121 TVDKTSEAAANNISKQTTTEADTDVDDSNAAANLQLEKLPNKVEIDGKYIYDNNKQVPT 180
 |||||

Qy 181 NFTLIADGKILHFDGTGAYTDTSDTVNKDIIVTTRSNLYKYNQVYDRAQSFHVDHYL 240
 |||||

Db 181 NFTLIADGKILHFDGTGAYTDTSDTVNKDIIVTTRSNLYKYNQVYDRAQSFHVDHYL 240
 |||||

Qy 241 TAEWYRKYILKDGKTWTQSTEDFRPLMTWPDQETQRYVYVYMAQLGINKTYDDT 300
 |||||

Db 241 TAEWYRKYILKDGKTWTQSTEDFRPLMTWPDQETQRYVYVYMAQLGINKTYDDT 300
 |||||

Qy 301 SNQLQNTAAATIAQIEAKITTLKNTDWLQRTISAFVKTQSAMNSDSEKPPDDHLQNGA 360
 |||||

Db 301 SNQLQNTAAATIAQIEAKITTLKNTDWLQRTISAFVKTQSAMNSDSEKPPDDHLQNGA 360
 |||||

Qy 361 VLYDNEGKLTYPYANSNYRILNRTPTNQTGKDPRTYADNTIGGYEFLANDVDSNPVVQ 420
 |||||

Db 361 VLYDNEGKLTYPYANSNYRILNRTPTNQTGKDPRTYADNTIGGYEFLANDVDSNPVVQ 420
 |||||

Qy 421 AEQLNWLHFLMNFNIYANDPDANFDSIRVDANVDADLLQIAGDYLKAAGIHKNDKA 480
 |||||

Db 421 AEQLNWLHFLMNFNIYANDPDANFDSIRVDANVDADLLQIAGDYLKAAGIHKNDKA 480
 |||||

Qy 481 ANDHLSILEAWSNDTPTYLHDDGDNMINMDNKLRLSLFLSLAKPLNQSGMPLTNSLV 540
 |||||

Db 481 ANDHLSILEAWSNDTPTYLHDDGDNMINMDNKLRLSLFLSLAKPLNQSGMPLTNSLV 540
 |||||

Qy 541 NRTDDNATAVPSPSIFIRAHDSVQDLIADIILKAEINPNVYVGYFTWEEIKKAEIYNK 600
 |||||

Db 541 NRTDDNATAVPSPSIFIRAHDSVQDLIADIILKAEINPNVYVGYFTWEEIKKAEIYNK 600
 |||||

Qy 601 DLLATEKKTHTYNALSVALLLTNKSSVPRVYVGYDMFTDDGOYMAHKTINYEAIETLLKA 660
 |||||

Db 601 DLLATEKKTHTYNALSVALLLTNKSSVPRVYVGYDMFTDDGOYMAHKTINYEAIETLLKA 660
 |||||

Qy 661 RIKYVSGQAMRNQOVGNSEIITSVRYGKGALKATDTGDRTRTTSVAVIEGNPNSLRLK 720
 |||||

Db 661 RIKYVSGQAMRNQOVGNSEIITSVRYGKGALKATDTGDRTRTTSVAVIEGNPNSLRLK 720
 |||||

QY 1 MDKKVYKURKKVWTVSVASAVMTLTTLSGGLVKADSNESKSOISNDSNTSVVTANE 60
DB 1 MDKKVYKLRKVKRWTVSVASAVMTLTTLSGGLVKADSNESKSOISNDSNTSVVTANE 60
QY 61 ESNVITEATSKOBAASQTNHTVTTSSSTSVVNPKVSVNPNYTVGETASNGEKLQNTT 120
DB 61 ESNVITEATSKOBAASQTNHTVTTSSSTSVVNPKVSVNPNYTVGETASNGEKLQNTT 120
QY 121 TVDKTSEAAANNISKOTTEADTDVIDDSNAANIQLILEKLPNVKEIDGKYIYDNNCKVRT 180
DB 121 TVDKTSEAAANNISKOTTEADTDVIDDSNAANIQLILEKLPNVKEIDGKYIYDNNCKVRT 180
QY 181 NFTLIADGKILHDETGAYTDSIDTVNKDIDVTRSNLYKKYNQVYDRSAQSEPHVDHYL 240
DB 181 NFTLIADGKILHDETGAYTDSIDTVNKDIDVTRSNLYKKYNQVYDRSAQSEPHVDHYL 240
QY 241 TAEWYRPKYILKDGKTWTQSTEKDFRPLLLMTWMPDQETQORQYVYTNMAQLGINKTYDDT 300
DB 241 TAEWYRPKYILKDGKTWTQSTEKDFRPLLLMTWMPDQETQORQYVYTNMAQLGINKTYDDT 300
QY 301 SNQQLNIAAATIQAIEAKITTLKNTDMLROTISAFVKTOSAWNSDSEKPPDDHLONGA 360
DB 301 SNQQLNIAAATIQAIEAKITTLKNTDMLROTISAFVKTOSAWNSDSEKPPDDHLONGA 360
QY 361 VLYDNEGKLTYPVANSNYRIILNFTPTNQTGKDPRTADNTIGGYEFLLANDVNSNPVQ 420
DB 361 VLYDNEGKLTYPVANSNYRIILNFTPTNQTGKDPRTADNTIGGYEFLLANDVNSNPVQ 420
QY 421 ASQNLWLHFLMFGNLYANDPANFDSIRVDAVDNVDADLLQIAGDYLKAAGKIHKNDKA 480
DB 421 ASQNLWLHFLMFGNLYANDPANFDSIRVDAVDNVDADLLQIAGDYLKAAGKIHKNDKA 480
QY 481 ANDHLSILEAWSNDPTPYLHDDGDNMINMDNKLRLSLLFSLAKPLNQRSGMPLIINSIV 540
DB 481 ANDHLSILEAWSNDPTPYLHDDGDNMINMDNKLRLSLLFSLAKPLNQRSGMPLIINSIV 540
QY 541 NRTDQNAETAAPVSFIRAHSEVQDLIADIKAEBINPNVVGYSFTMBEIIKKAPEIYNK 600
DB 541 NRTDQNAETAAPVSFIRAHSEVQDLIADIKAEBINPNVVGYSFTMBEIIKKAPEIYNK 600
QY 601 DLLATEKKYTHNTALSYALLTNKSSVPRVYGDMTDQDQGYMAHKTINYEAIETLLKA 660
DB 601 DLLATEKKYTHNTALSYALLTNKSSVPRVYGDMTDQDQGYMAHKTINYEAIETLLKA 660
QY 661 RIKYVSGGQMRNQVGNSEIITSVRYGKALKATDGTDRTRTSVAVIEGNPSLRK 720
DB 661 RIKYVSGGQMRNQVGNSEIITSVRYGKALKATDGTDRTRTSVAVIEGNPSLRK 720
QY 721 ASDRVVNVNNGAAHKNQAYRPLLLTTDNGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKG 780
DB 721 ASDRVVNVNNGAAHKNQAYRPLLLTTDNGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKG 780
QY 781 YANPQVSGYLGWVPVGAALIIRKMFALRLARPHQOMASVHQNAAALDSRVNFEQFSNFPAPA 840
DB 781 YANPQVSGYLGWVPVGAALIIRKMFALRLARPHQOMASVHQNAAALDSRVNFEQFSNFPAPA 840
QY 841 TKKEEYTNVIAKNVDKFAEWGVTDFEMAPQVSVSDGSLDSVLQNGVAFTRDYLGIS 900
DB 841 TKKEEYTNVIAKNVDKFAEWGVTDFEMAPQVSVSDGSLDSVLQNGVAFTRDYLGIS 900
QY 901 KPNKYGTADDLKAIALKHSKGIKVMADVPDQMYAFPEKEVVVTRVVDKYGTPVAGSQI 960
DB 901 KPNKYGTADDLKAIALKHSKGIKVMADVPDQMYAFPEKEVVVTRVVDKYGTPVAGSQI 960
QY 961 KNTLYVVDGKSGKQQAQYGAFFELQAKYPELFARKQISTGVPMPSVKIKOWSAKY 1020
DB 961 KNTLYVVDGKSGKQQAQYGAFFELQAKYPELFARKQISTGVPMPSVKIKOWSAKY 1020
QY 1021 FNGTNILGRGAGYVLKQDQATNTYFNISDNKEINFLPKTLNODSQVGSYDGKGVYVYST 1080
DB 1021 FNGTNILGRGAGYVLKQDQATNTYFNISDNKEINFLPKTLNODSQVGSYDGKGVYVYST 1080
QY 1081 SGYQAKNTFISEGDKWYYPDNNGYMVTGAQSIGNGVNYFSLNGLQLRDAILKNEDGTAY 1140

DB 1081 SGYQAKNTFISEGDKWYYPDNNGYMVTGAQSIGNGVNYFSLNGLQLRDAILKNEDGTAY 1140
QY 1141 YGNDGRRYENGYYQFMGSGVWRHPNNGEMSVGLTVIDGOVQYFDEMGGYQAKGKFTVTADGK 1200
DB 1141 YGNDGRRYENGYYQFMGSGVWRHPNNGEMSVGLTVIDGOVQYFDEMGGYQAKGKFTVTADGK 1200
QY 1201 IRYFDKQSGNMVNRNRIENEEGKWLVLGEDGAAVTGSQTINGQHLYFRANGVQVKGFEV 1260
DB 1201 IRYFDKQSGNMVNRNRIENEEGKWLVLGEDGAAVTGSQTINGQHLYFRANGVQVKGFEV 1260
QY 1261 DHHGRISYYDNGSGDOIIRNRFVRNAQOMFYFDNNGYAVTGARTINGQLLYFRANGVQV 1320
DB 1261 DHHGRISYYDNGSGDOIIRNRFVRNAQOMFYFDNNGYAVTGARTINGQLLYFRANGVQV 1320
QY 1321 GEFVTDTRYGRISYYDNGSGDOIIRNRFVRNAQOMFYFDNNGYAVTGARTINGQHLYFRAN 1380
DB 1321 GEFVTDTRYGRISYYDNGSGDOIIRNRFVRNAQOMFYFDNNGYAVTGARTINGQHLYFRAN 1380
QY 1381 GVQVKGEFVTDHGRISYYDNGSGDOIIRNRFVRNAQOMFYFDNNGYAVTGARTINGOHL 1440
DB 1381 GVQVKGEFVTDHGRISYYDNGSGDOIIRNRFVRNAQOMFYFDNNGYAVTGARTINGOHL 1440
QY 1441 YFRANGVQVKGFEVTDTRYGRISYYDANSGERVRAIN 1475
DB 1441 YFRANGVQVKGFEVTDTRYGRISYYDANSGERVRAIN 1475
RESULT 9
AAU98035
ID AAU98035 standard; protein; 1475 AA.
XX
AC AAU98035;
XX
DT 27-AUG-2002 (first entry)
XX
DE S. mutans glucosyltransferase GTFB mutant D457N/D571K.
XX
KW Glucosyltransferase; GTPB; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture; mutant; mutein.
XX
OS Streptococcus mutans.
OS Synthetic.
XX
XX Key Location/Qualifiers
FH Misc-difference 457
FT Misc-difference 571 /note= "Wild-type Asp substituted by Asn"
FT Misc-difference 571 /note= "Wild-type Asp substituted by Lys"
XX
PN US2002031826-A1.
XX
PD 14-MAR-2002.
XX
XX 19-DEC-2000; 2000US-00740274.
XX
XX 07-JUN-1995; 95US-00478704.
PR 07-JUN-1995; 95US-00482711.
PR 07-JUN-1995; 95US-00485243.
PR 16-JAN-1998; 98US-00007999.
PR 16-JAN-1998; 98US-00008172.
PR 20-JAN-1998; 98US-00009620.
PR 11-DEC-1998; 98US-00210361.
XX
XX (NICH/) NICHOLS S E.
XX
XX Nichols SE;
XX
XX WPI; 2002-414332/44.
XX
PT Glucosyltransferase B or D protein useful for producing a glucan useful
PT as substitutes for and additions to modified starch and latexes in paper

Qy	1	MDKKVRYKLRKVKKGWVTVSVASAVMTTTLTSSGGLVKADSNESKQISNDSENTSVVTANE	60
Db	1	MDKKVRYKLRKVKKGWVTVSVASAVMTTTLTSSGGLVKADSNESKQISNDSENTSVVTANE	60
Qy	61	ESNVITEATSKEOEAASSQTNHTVTSSSSTSVNPKVVSNPYTVGETASNGEKLQNTT	120
Db	61	ESNVITEATSKEOEAASSQTNHTVTSSSSTSVNPKVVSNPYTVGETASNGEKLQNTT	120
Qy	121	TVDKTSAAANNISKQTTTEADTVDDNSAANLQLEKLPNVKEIDGKYIYDNNKGVRT	180
Db	121	TVDKTSAAANNISKQTTTEADTVDDNSAANLQLEKLPNVKEIDGKYIYDNNKGVRT	180
Qy	181	NFTLIADGKILHFDGTGAYTDTSDITVKNKJIVTTRSNLYKKYNOVYDRSAQSFEHVDHYL	240
Db	181	NFTLIADGKILHFDGTGAYTDTSDITVKNKJIVTTRSNLYKKYNOVYDRSAQSFEHVDHYL	240
Qy	241	TAESWYRPKYILKQKQWTQSTEKDFRPLLMTWPDQETORQYVYNNNAQJINGKTYDDT	300
Db	241	TAESWYRPKYILKQKQWTQSTEKDFRPLLMTWPDQETORQYVYNNNAQJINGKTYDDT	300
Qy	301	SNQJLNIAAATIOAKIEAKITTLKNTDMLRQTTISAFVKTSQSAWNSDSEKPFDDHLONGA	360
Db	301	SNQJLNIAAATIOAKIEAKITTLKNTDMLRQTTISAFVKTSQSAWNSDSEKPFDDHLONGA	360
Qy	361	VLYNEGKJTPYANSNRYILNRTPTNQTKKDPRTYADNTTGGYEFLLANDVDNSNPVVQ	420
Db	361	VLYNEGKJTPYANSNRYILNRTPTNQTKKDPRTYADNTTGGYEFLLANDVDNSNPVVQ	420
Qy	421	AEOLNMLHFLNFGNIVANDPDANFDSIRVDADVNDADILLOIAGDYLKAAGKHKNKA	480
Db	421	AEOLNMLHFLNFGNIVANDPDANFDSIRVDADVNDADILLOIAGDYLKAAGKHKNKA	480
Qy	481	ANDHLSILLEANSDDNTPYLHDDGDNMINMDKNLRLSLLFSLAKPLNORSGNWPLITNSLV	540
Db	481	ANDHLSILLEANSDDNTPYLHDDGDNMINMDKNLRLSLLFSLAKPLNORSGNWPLITNSLV	540
Qy	541	NRTDDNAETAAPVSPYSFIRAHDSSEVQTLIAKIIKAEINPNVVGYSFTWEEIKKAFETYNK	600
Db	541	NRTDDNAETAAPVSPYSFIRAHDSSEVQTLIAKIIKAEINPNVVGYSFTWEEIKKAFETYNK	600
Qy	601	DLLEATEKYTHYNTALSYALLLTKNSVPRVYVYGDMTDDGOYMAHKTINYEAIETLLKA	660
Db	601	DLLEATEKYTHYNTALSYALLLTKNSVPRVYVYGDMTDDGOYMAHKTINYEAIETLLKA	660
Qy	661	RIKVVSGGOAMRNQOVGNSEIITSVRVYKGGALKATDGTDRTRTSGVAVLEGNNPSURLK	720
Db	661	RIKVVSGGOAMRNQOVGNSEIITSVRVYKGGALKATDGTDRTRTSGVAVLEGNNPSURLK	720
Qy	721	ASDRVVMNGAAHKNQAYRPLLLTTDNGIKAYHSDQEAAGLVRYTNDRGELIIFTAADIKG	780
Db	721	ASDRVVMNGAAHKNQAYRPLLLTTDNGIKAYHSDQEAAGLVRYTNDRGELIIFTAADIKG	780
Qy	781	YANPQVSGYLGWVPVPGAALIKMFALRLARPHQOMASVHONAAALDSRWMEGSGNFOAFA	840
Db	781	YANPQVSGYLGWVPVPGAALIKMFALRLARPHQOMASVHONAAALDSRWMEGSGNFOAFA	840
Qy	841	TKKSEYTNVVIATKNDVKAFAEWGVTDFEMAPQVYSSDTGSLDSVIQNGYAFTRYDILGIS	900
Db	841	TKKSEYTNVVIATKNDVKAFAEWGVTDFEMAPQVYSSDTGSLDSVIQNGYAFTRYDILGIS	900
Qy	901	KPNKYGTADDLVAIKALHSGIKVMADWPDQMYAPPEKEVVTATRVDKYGTVPAGSQI	960
Db	901	KPNKYGTADDLVAIKALHSGIKVMADWPDQMYAPPEKEVVTATRVDKYGTVPAGSQI	960
Qy	961	KNTLYVVDGKSGXKQQAQYGGAFLEBLQAKYPELFARKQISTGVPMDPVSKIKQMSAKY	1020
Db	961	KNTLYVVDGKSGXKQQAQYGGAFLEBLQAKYPELFARKQISTGVPMDPVSKIKQMSAKY	1020
Qy	1021	FNGTNIILGRGAGYVLKQOATNTYFNIISDNKEINFLPKTLLNODSOVGFSDGKGYVYYST	1080
Db	1021	FNGTNIILGRGAGYVLKQOATNTYFNIISDNKEINFLPKTLLNODSOVGFSDGKGYVYYST	1080

QY	1081	SGYQAKNTFTISEGDKWYTFDNNNGYMTGAOSINGVNYFYFLSNGIQLRDAILKKNEDGTYAY	1144
Db	1081	SGYQAKNTFTISEGDKWYTFDNNNGYMTGAOSINGVNYFYFLSNGIQLRDAILKKNEDGTYAY	1140
QY	1141	YGNDRRYENGYQFMGSVWRHFNNGRMSVGLTVIDGQVOYFDEMGSYQAKGKFVTTADGK	1200
Db	1141	YGNDRRYENGYQFMGSVWRHFNNGRMSVGLTVIDGQVOYFDEMGSYQAKGKFVTTADGK	1200
QY	1201	IRYFDKQSGNMRYNRFTENEEGKWLILGEDGAAVTGSQTINGOHLYFRANGVQVKGFEV	1260
Db	1201	IRYFDKQSGNMRYNRFTENEEGKWLILGEDGAAVTGSQTINGOHLYFRANGVQVKGFEV	1260
QY	1261	DHGRISYYDGNSGDQIRNRFVRNAQQWYFDNNGYAVTGARTINGQLLYFRANGVQVK	1320
Db	1261	DHGRISYYDGNSGDQIRNRFVRNAQQWYFDNNGYAVTGARTINGQLLYFRANGVQVK	1320
QY	1321	GEFVTDYRGRISYYDGNSGDQIRNRFVRNAQQWYFDNNGYAVTGARTINGOHLYFRAN	1380
Db	1321	GEFVTDYRGRISYYDGNSGDQIRNRFVRNAQQWYFDNNGYAVTGARTINGOHLYFRAN	1380
QY	1381	GVQVKGFEVTDHGRISYYDGNSGDQIRNRFVRNAQQWYFDNNGYAVTGARTINGOHL	1440
Db	1381	GVQVKGFEVTDHGRISYYDGNSGDQIRNRFVRNAQQWYFDNNGYAVTGARTINGOHL	1440
QY	1441	YFRANGVQVKGFEVTDYRGRISYYDANSRGRVIRIN	1475
Db	1441	YFRANGVQVKGFEVTDYRGRISYYDANSRGRVIRIN	1475
RESULT 12			
AAU98037			
ID	AAU98037 standard; protein; 1475 AA.		
AC	AAU98037;		
XX			
DT	27-AUG-2002 (first entry)		
XX			
DE	S. mutans glucosyltransferase GTFB mutant D567T/D571K/K1014T.		
XX			
KW	Glucosyltransferase; GTFB; transgenic plant; paper sizing;		
KW	coating composition; glucan; starch; latex; thermoplastic molecule;		
KW	amyloplast; vacuole; paper manufacture; mutant; mutein.		
XX			
OS	Streptococcus mutans.		
OS	Synthetic.		
XX			
EH	Key Location/Qualifiers		
FT	Misc-difference 567	/note= "Wild-type Asp substituted by Thr"	
FT	Misc-difference 571	/note= "Wild-type Asp substituted by Lys"	
FT	Misc-difference 1014	/note= "Wild-type Lys substituted by Thr"	
XX			
PN	US2002031826-A1.		
XX			
PD	14-MAR-2002.		
XX			
PF	19-DEC-2000; 2000US-00740274.		
XX			
PR	07-JUN-1995;	95US-00478704.	
PR	07-JUN-1995;	95US-00482711.	
PR	07-JUN-1995;	95US-00485243.	
PR	16-JAN-1998;	98US-00007999.	
PR	16-JAN-1998;	98US-00008172.	
PR	20-JAN-1998;	98US-00009620.	
XX	11-DEC-1998;	98US-00210361.	
XX	(NICH/) NICHOLS S E.		
PA			
XX			
FI	Nichols SE;		
XX			
DR	WPI; 2002-414332/44.		

XX Glucosyltransferase B or D protein useful for producing a glucan useful
PT as substitutes for and additions to modified starch and latexes in paper
PT manufacture, comprises mutations in specific positions.
XX
PS
PS Claim 36; Page; 44pp; English.
XX
CC The invention an isolated protein comprising a glucosyltransferase (GTF)
CC B polypeptide having changes at position from 1448V, D457N, D567T,
CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/K1014T,
CC I448V/D457N/D567T/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
CC complementary polynucleotide, a ribonucleic acid sequence encoding the
CC GTF mutant, an expression cassette comprising the polynucleotide operably
CC linked to a promoter, a vector comprising the expression cassette, host
CC cell introduced with the vector, a transgenic plant comprising the
CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
CC coating composition comprising a glucan produced in a plant transformed
CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
CC thermoplastic molecule or their combinations or glucan and starch where
CC the glucan is produced in the amyloplast and/or vacuole or a maize line
CC deficient in starch biosynthesis, transformed with a gene encoding a
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
CC comprising the glucan (paper sizing/coating agent). The vector is useful
CC for producing a glucan in a plant. The method comprises transforming a
CC plant cell with the vector, growing the plant cell under plant growing
CC conditions to produce a regenerated plant and inducing expression of the
CC polynucleotide for a time sufficient to produce the glucan in the
CC regenerated plant, where the vector contains a transit sequence from
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
CC is produced in the amyloplast of potato or the vacuole of sugar beet.
CC Glucans are useful as substitutes for and additions to modified starch
CC and latexes in paper manufacture. Unlike prior art techniques, which
CC require input materials that produce chemical effluents, paper
CC manufacture utilises the glucan produced by GTF, which utilises
CC biologically produced input materials, is more cost-effective and
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
CC properties and impart gloss to the paper during coating step. The present
CC sequence represents a GTFB mutant of the invention. Note: The present
CC sequence is not shown in the specification but was created by the indexer
CC using the GTFB sequence appearing as AAU98027 and the information in
CC claim 36
XX
SQ Sequence 1475 AA;

Query Match 99.7%; Score 7721; DB 5; Length 1475;
Best Local Similarity 99.8%; Fred. No. 0;
Matches 1472; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MDKVKRYKLRKVKRWTVSVASAVMTLTTLTSGGLVKADSNESKQISNDNSNTSVVTANE 60
DB 1 MDKVKRYKLRKVKRWTVSVASAVMTLTTLTSGGLVKADSNESKQISNDNSNTSVVTANE 60
QY 61 ESNVITEATSKQEAASSQTNHTVTTSSSTSVVNPKEVSNPYTVGTASNGEKLQNTT 120
DB 61 ESNVITEATSKQEAASSQTNHTVTTSSSTSVVNPKEVSNPYTVGTASNGEKLQNTT 120
QY 121 TVDKTSEAAANNISKQTTTADTDVDDSNANLQILEKLPNVKEIDGKYYYDNNKVRT 180
DB 121 TVDKTSEAAANNISKQTTTADTDVDDSNANLQILEKLPNVKEIDGKYYYDNNKVRT 180
QY 181 NFTLIADGKILHFDSTGAYTDTSDTVNKDIDVTTTSNLYKKYNQVYDRSAQSFHVHDYL 240
DB 181 NFTLIADGKILHFDSTGAYTDTSDTVNKDIDVTTTSNLYKKYNQVYDRSAQSFHVHDYL 240
QY 241 TAEWYRPKYILKDGKWTQSTEKDFRPLMTWPDQETQRQYVNMNAQLGINKTYDDT 300
DB 241 TAEWYRPKYILKDGKWTQSTEKDFRPLMTWPDQETQRQYVNMNAQLGINKTYDDT 300
QY 301 SNQLQLNTAATIQAIEAKITTLKNTDWLQRTISAFVKTOSAWNSDSEKPPDDHLQNGA 360

DB 301 SNQLQLNTAATIQAIEAKITTLKNTDWLQRTISAFVKTOSAWNSDSEKPPDDHLQNGA 360
QY 361 VLYDNEGKLTTPYANSNYRILNRTPTNQTGKDPRTYADNTTIGGYEFLANDVDSNPVVQ 420
DB 361 VLYDNEGKLTTPYANSNYRILNRTPTNQTGKDPRTYADNTTIGGYEFLANDVDSNPVVQ 420
QY 421 AEQLNLHFLMNFNIYANDPDANPDSIRVDAVNDVADADLQIAGDYLLKAAKGIHNKDA 480
DB 421 AEQLNLHFLMNFNIYANDPDANPDSIRVDAVNDVADADLQIAGDYLLKAAKGIHNKDA 480
QY 481 ANDHLSIILEAMSDNDTPYLHDDGDNMIMNDKRLSLLSFLAKPLNQSGMNPPLTNSLV 540
DB 481 ANDHLSIILEAMSDNDTPYLHDDGDNMIMNDKRLSLLSFLAKPLNQSGMNPPLTNSLV 540
QY 541 NRTDDNAETAAPVPSYFIRAHDSVQDLIADIIKAEINPNVVGYSFTMEEIKKAFEIYNK 600
DB 541 NRTDDNAETAAPVPSYFIRAHDSVQDLIADIIKAEINPNVVGYSFTMEEIKKAFEIYNK 600
QY 601 DLLATEKKYTHYNTALSYALLLTNNKSSVPRVYVYGMFTDDGQYMAHKTINYBAIETLLKA 660
DB 601 DLLATEKKYTHYNTALSYALLLTNNKSSVPRVYVYGMFTDDGQYMAHKTINYBAIETLLKA 660
QY 661 RIKYVSGQAMRNQOVGNSEIITSVRYCGKGAUKATDGTGRTTTSVAVIEGNNPSLRUK 720
DB 661 RIKYVSGQAMRNQOVGNSEIITSVRYCGKGAUKATDGTGRTTTSVAVIEGNNPSLRUK 720
QY 721 ASDRVVNMGAHKNQAYRPLLLTTDNGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKG 780
DB 721 ASDRVVNMGAHKNQAYRPLLLTTDNGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKG 780
QY 781 YANPQVSYGLGVVPVGAALIKMFALRLARPHQQWASVHQAALDSRVMFEGSFNQAPA 840
DB 781 YANPQVSYGLGVVPVGAALIKMFALRLARPHQQWASVHQAALDSRVMFEGSFNQAPA 840
QY 841 TKKEEYTNVITAKNVDFEAEWGVTDFFEMAPQVSVSTGDSFLDSVTLQNGVAFYFTRDVLGTS 900
DB 841 TKKEEYTNVITAKNVDFEAEWGVTDFFEMAPQVSVSTGDSFLDSVTLQNGVAFYFTRDVLGTS 900
QY 901 KPNKYGTADDLVKAIKALHSKGIKVMADWPDQWYAFPEKEVVTATRVKDTPTVAGSQI 960
DB 901 KPNKYGTADDLVKAIKALHSKGIKVMADWPDQWYAFPEKEVVTATRVKDTPTVAGSQI 960
QY 961 KNTLYVVDGKSSGKQQAKYCGAFLEELQAKYPELFARKQISTGVMPDPSVKIKOWSAKY 1020
DB 961 KNTLYVVDGKSSGKQQAKYCGAFLEELQAKYPELFARKQISTGVMPDPSVKIKOWSAKY 1020
QY 1021 FNGTNILGRGAGYVVKDQATNTYFNI SDNKEINFLPKTLNODSVOGFSYDGKGYYVYST 1080
DB 1021 FNGTNILGRGAGYVVKDQATNTYFNI SDNKEINFLPKTLNODSVOGFSYDGKGYYVYST 1080
QY 1081 SGYQAKNTFISEGDKWYTFDNNGYMVTGAQOSINGVNYVYFLSNGQLRLDAILKNEGTAY 1140
DB 1081 SGYQAKNTFISEGDKWYTFDNNGYMVTGAQOSINGVNYVYFLSNGQLRLDAILKNEGTAY 1140
QY 1141 YGNDGRRYENGYYQFMGSMWRHFNNGEMSVGLTVIDGQVQYFDEMGYQAKGKFTVTDGK 1200
DB 1141 YGNDGRRYENGYYQFMGSMWRHFNNGEMSVGLTVIDGQVQYFDEMGYQAKGKFTVTDGK 1200
QY 1201 IRYFDKSGNNMYRNFRTENEESKWLILGEDGAAVTGSOTINGOHLYFRANGVQKGEFVT 1260
DB 1201 IRYFDKSGNNMYRNFRTENEESKWLILGEDGAAVTGSOTINGOHLYFRANGVQKGEFVT 1260
QY 1261 DHGRIISYDNGSGDQIRNRFVRNAQOGWFFYFNNNGYAVTGARTINGQLLYFRANGVQVK 1320
DB 1261 DHGRIISYDNGSGDQIRNRFVRNAQOGWFFYFNNNGYAVTGARTINGQLLYFRANGVQVK 1320
QY 1321 GEFVTDYGRISYYDNGSGDQIRNRFVRNAQOGWFFYFNNNGYAVTGARTINGOHLYFRAN 1380
DB 1321 GEFVTDYGRISYYDNGSGDQIRNRFVRNAQOGWFFYFNNNGYAVTGARTINGOHLYFRAN 1380
QY 1381 GVQVKGFEVTDYGRISYYDNGSGDQIRNRFVRNAQOGWFFYFNNNGYAVTGARTINGOHL 1440

Db 1381 GVQVKGFEVTDHGRISYYDNGSGDQIRNFVRNAGQWTFYDNNGVAVTGARTINGQHL 1440

QY 1441 YFRANGVQVKGFEVTDHGRISYYDANGSERVIN 1475
 |||||

Db 1441 YFRANGVQVKGFEVTDHGRISYYDANGSERVIN 1475
 |||||

RESULT 13

ID AAU98039

AC AAU98039 standard; protein; 1475 AA.

XX AAU98039;

XX

XX 27-AUG-2002 (first entry)

XX

XX S. mutans glucosyltransferase GTFB mutant YYY169-171AAA.

DE

XX Glucosyltransferase; GTFB; transgenic plant; paper sizing;

KW coating composition; glucan; starch; latex; thermoplastic molecule;

KW amyloplast; vacuole; paper manufacture; mutant; mutein.

XX

OS Streptococcus mutans.

OS Synthetic.

XX

XX

XX Key Location/Qualifiers

FT Misc-difference 169..171

FT /note= "Wild-type Tyr-Tyr substituted by Ala-Ala"

XX

XX US2002031826-A1.

XX

XX 14-MAR-2002.

XX

XX 19-DEC-2000; 2000US-00740274.

XX

XX 07-JUN-1995; 95US-00478704.

PR 07-JUN-1995; 95US-00482711.

PR 07-JUN-1995; 95US-00485243.

PR 16-JAN-1998; 98US-00007999.

PR 16-JAN-1998; 98US-00008172.

PR 20-JAN-1998; 98US-00009620.

PR 11-DEC-1998; 98US-00210361.

XX

XX (NICH/) NICHOLS S E.

XX

XX Nichols SE;

XX

XX WPI; 2002-414332/44.

XX

XX Glucosyltransferase B or D protein useful for producing a glucan useful

PT as substitutes for and additions to modified starch and latexes in paper

PT manufacture, comprises mutations in specific positions.

XX

XX Claim 36; Page: 44pp; English.

XX

XX The invention an isolated protein comprising a glucosyltransferase (GTF)

CC B polypeptide having changes at position from 1448V, D457N, D567T,

CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T,

CC 1448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a

CC GTF D polypeptide having changes at positions from 1589D, T589E, N471D,

CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the

CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its

CC complementary polynucleotide, a ribonucleic acid sequence encoding the

CC GTF mutant, an expression cassette comprising the polynucleotide operably

CC linked to a promoter, a vector comprising the expression cassette, host

CC cell introduced with the vector, a transgenic plant comprising the

CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or

CC coating composition comprising a glucan produced in a plant transformed

CC with a gene encoding the mutant GTF, wild type or, starch, a latex,

CC thermoplastic molecule or their combinations or glucan and starch where

CC the glucan is produced in the amyloplast and/or vacuole or a maize line

CC deficient in starch biosynthesis, transformed with a gene encoding a

CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper

CC comprising the glucan (paper sizing/coating agent). The vector is useful

CC for producing a glucan in a plant. The method comprises transforming a

CC plant cell with the vector, growing the plant cell under plant growing

CC conditions to produce a regenerated plant and inducing expression of the

CC polynucleotide for a time sufficient to produce the glucan in the

CC regenerated plant, where the vector contains a transit sequence from

CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and

CC chlorophyll AB binding protein to produce a transgenic plant, and glucan

CC is produced in the amyloplast of potato or the vacuole of sugar beet.

CC Glucans are useful as substitutes for and additions to modified starch

CC and latexes in paper manufacture. Unlike prior art techniques, which

CC require input materials that produce chemical effluents, paper

CC manufacture utilising the glucan produced by GTF, which utilises

CC biologically produced input materials, is more cost-effective and

CC environmentally friendly. Moreover, glucans also exhibit thermoplastic

CC properties and impart gloss to the paper during coating step. The present

CC sequence represents a GTFB mutant of the invention. Note: The present

CC sequence is not shown in the specification but was created by the indexer

CC using the GTFB sequence appearing as AAU98027 and the information in

CC claim 36

XX

SQ Sequence 1475 AA;

Query Match 99.7%; Score 7714; DB 5; Length 1475;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 1472; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MDKVRVYKLRVKRQWTVSVASAVMTLTLSGGLVKADSNESKSOISNDSNTSVVTANE 60
 |||||

Db 1 MDKVRVYKLRVKRQWTVSVASAVMTLTLSGGLVKADSNESKSOISNDSNTSVVTANE 60
 |||||

QY 61 ESNVITEATSKQEAASSQTNHTVTSSSTSVVNPKEVSNPYTVGTETASNGEKLQNTT 120
 |||||

Db 61 ESNVITEATSKQEAASSQTNHTVTSSSTSVVNPKEVSNPYTVGTETASNGEKLQNTT 120
 |||||

QY 121 TVDKTSEAAANNISKQTTTADTDVDDSNAAANLQILEKLPNVKEIDGKYYYDNNGKVRT 180
 |||||

Db 121 TVDKTSEAAANNISKQTTTADTDVDDSNAAANLQILEKLPNVKEIDGKYYYDNNGKVRT 180
 |||||

QY 181 NFTLIADGKILHFDGTGAYTDSITVANKDITVTSNLYKKYNQVYDRSAQSPHEVDHYL 240
 |||||

Db 181 NFTLIADGKILHFDGTGAYTDSITVANKDITVTSNLYKKYNQVYDRSAQSPHEVDHYL 240
 |||||

QY 241 TAESWYRPKYILKDGKWTQSTKDFRPLLLMTWPDQETQRYVNTMNAQLGINKTYDDT 300
 |||||

Db 241 TAESWYRPKYILKDGKWTQSTKDFRPLLLMTWPDQETQRYVNTMNAQLGINKTYDDT 300
 |||||

QY 301 SNQLQNTAAATIAQIEAKITTLKNTDHLRTISAFVKTQSAWNSDSKPFDDHLLQNGA 360
 |||||

Db 301 SNQLQNTAAATIAQIEAKITTLKNTDHLRTISAFVKTQSAWNSDSKPFDDHLLQNGA 360
 |||||

QY 361 VLYDNEGKLTVPYANSNYRILNRTPTNQTGKDPRTYADNTTIGSYEFLLANDVDSNPVQ 420
 |||||

Db 361 VLYDNEGKLTVPYANSNYRILNRTPTNQTGKDPRTYADNTTIGSYEFLLANDVDSNPVQ 420
 |||||

QY 421 AEQLNLWHLFMNFGNIYANDPDANFDSIRVDADVNDVADLLQIAGDYLAKAAGIHKNDKA 480
 |||||

Db 421 AEQLNLWHLFMNFGNIYANDPDANFDSIRVDADVNDVADLLQIAGDYLAKAAGIHKNDKA 480
 |||||

QY 481 ANDHLSILEAWSDNDTPYLHDDGDNINMNDKRLSLFLPSLAKPLNQRSGMPLITNSLV 540
 |||||

Db 481 ANDHLSILEAWSDNDTPYLHDDGDNINMNDKRLSLFLPSLAKPLNQRSGMPLITNSLV 540
 |||||

QY 541 NRTDINAETAAPVPSYFIRAHDSVQDITADIKAIEINPVVGYSTFMEIEIKAFIYNK 600
 |||||

Db 541 NRTDINAETAAPVPSYFIRAHDSVQDITADIKAIEINPVVGYSTFMEIEIKAFIYNK 600
 |||||

QY 601 DLLATEKYYTHNTALSYALLLTNKSVPVRYVYGMFTDDGQYMAHKTINYEAIETLLKA 660
 |||||

Db 601 DLLATEKYYTHNTALSYALLLTNKSVPVRYVYGMFTDDGQYMAHKTINYEAIETLLKA 660
 |||||

QY 661 RIKYVSGQAMRNQGVNSEIITSVRYGKALKATDGTDRTRTTSVAVIEGNPNPSRLRK 720
 |||||

Db 661 RIKYVSGQAMRNQGVNSEIITSVRYGKALKATDGTDRTRTTSVAVIEGNPNPSRLRK 720
 |||||

CC sequence is not shown in the specification but was created by the indexer
CC using the GTFB sequence appearing as AAU98027 and the information in
CC claim 36
XX
SQ Sequence 1475 AA;

Query Match 99.6%; Score 7711; DB 5; Length 1475;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1469; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy	1	MDKVRKYLKRVKRWVTVSVASAVMTLTLTSGGLVKADSNESKQISNDSTSVVTANE	60
Db	1	MDKVRKYLKRVKRWVTVSVASAVMTLTLTSGGLVKADSNESKQISNDSTSVVTANE	60
Qy	61	ENSVITEATSKQEAASSQTNHTVTSSSSSTSVVNPKVVSNTYVGTETASNGEKLQNTT	120
Db	61	ENSVITEATSKQEAASSQTNHTVTSSSSSTSVVNPKVVSNTYVGTETASNGEKLQNTT	120
Qy	121	TVDTKSEAAANNISKQTTADTDVDDSDNAANLQILEKLPNVEIDGKYKYDNNNGKVRT	180
Db	121	TVDTKSEAAANNISKQTTADTDVDDSDNAANLQILEKLPNVEIDGKYKYDNNNGKVRT	180
Qy	181	NFTLLADGKILHFDGTGATDTSIDTVNKDIIVTTRSNLYKKYNOYVDRSAQSFHVHVL	240
Db	181	NFTLLADGKILHFDGTGATDTSIDTVNKDIIVTTRSNLYKKYNOYVDRSAQSFHVHVL	240
Qy	241	TAESWYRPKYILKDGKTWTQSTKDFRPLLMTWPDQETQROYVYVMAAQLGINKTYDDT	300
Db	241	TAESWYRPKYILKDGKTWTQSTKDFRPLLMTWPDQETQROYVYVMAAQLGINKTYDDT	300
Qy	301	SNQLQNTAAATIAKIEAKITTLKNTDMLRTISAFVKTQSAMNSDSEKPPDDHLQNGA	360
Db	301	SNQLQNTAAATIAKIEAKITTLKNTDMLRTISAFVKTQSAMNSDSEKPPDDHLQNGA	360
Qy	361	VLYDNEGKLTPTVANSNYRLIARTPTNQTCKDPRYTAQNTICGYEFLANDVDSNPVQ	420
Db	361	VLYDNEGKLTPTVANSNYRLIARTPTNQTCKDPRYTAQNTICGYEFLANDVDSNPVQ	420
Qy	421	AEQLNLHFLMNFNIYANDPANFDSIRVDADVNDVADLLQIAGDYLKAAKGIHNKDKA	480
Db	421	AEQLNLHFLMNFNIYANDPANFDSIRVDADVNDVADLLQIAGDYLKAAKGIHNKDKA	480
Qy	481	ANDHLSILEAASNDTPYLHDDGDMNMDNKLRLSLFLSLAKPLNQRSGMPLTNSLV	540
Db	481	ANDHLSILEAASNDTPYLHDDGDMNMDNKLRLSLFLSLAKPLNQRSGMPLTNSLV	540
Qy	541	NRTDDNAETAAPVPSYFIRAHDSVQDIIADIKAEPNVVGVYFTWEEIKKAFEIYNK	600
Db	541	NRTDDNAETAAPVPSYFIRAHDSVQDIIADIKAEPNVVGVYFTWEEIKKAFEIYNK	600
Qy	601	DLLATEKKYTHYNTALSVALLLTNKSSVPRVYVYGMFTDDGOYMAHKTINYEAIETLKA	660
Db	601	DLLATEKKYTHYNTALSVALLLTNKSSVPRVYVYGMFTDDGOYMAHKTINYEAIETLKA	660
Qy	661	RIKYVSGQAMRNQGVNSEIITSVRYGKALKATDTGDRTRTSVAVIEGNNPSRLK	720
Db	661	RIKYVSGQAMRNQGVNSEIITSVRYGKALKATDTGDRTRTSVAVIEGNNPSRLK	720
Qy	721	ASDRVVVNGAAHKQAYRPLLLTTDNGIKAYHSQDEAAGLVRYTNDRGELIFTAADIKG	780
Db	721	ASDRVVVNGAAHKQAYRPLLLTTDNGIKAYHSQDEAAGLVRYTNDRGELIFTAADIKG	780
Qy	781	YANPOVSGVLGVVPGVGAALIKMPALRLARPHQOMASVHQAALDSRVMEFGFSFOAPA	840
Db	781	YANPOVSGVLGVVPGVGAALIKMPALRLARPHQOMASVHQAALDSRVMEFGFSFOAPA	840
Qy	841	TKXEYTNVVIKXNDKFAEAGVTDTFEMAPQVVSSTGDSFLDSVITQNGYAFTRDYLGIS	900
Db	841	TKXEYTNVVIKXNDKFAEAGVTDTFEMAPQVVSSTGDSFLDSVITQNGYAFTRDYLGIS	900
Qy	901	KPNKYGTADDLVKAIKALHSGIKVMADWVPDMYAFPEKEVVTATRVKDTGTPVAGSQI	960
Db	901	KPNKYGTADDLVKAIKALHSGIKVMADWVPDMYAFPEKEVVTATRVKDTGTPVAGSQI	960

Qy	961	KNTLYVVDGSSGKQQAQYGGAFLEELQAKYPELPFARKQISTGVPMDSVKIKOWSAKY	1020
Db	961	KNTLYVVDGSSGKQQAQYGGAFLEELQAKYPELPFARKQISTGVPMDSVKIKOWSAKY	1020
Qy	1021	FNGTNILGRGAGYVLKQQAATNTYFNISDNKEINFLPKTLNLDSDSQVGSYDGKGYVYST	1080
Db	1021	FNGTNILGRGAGYVLKQQAATNTYFNISDNKEINFLPKTLNLDSDSQVGSYDGKGYVYST	1080
Qy	1081	SGYQAKNTFISGDKWYFDDNNGYVMTGAQSTNGVNYFSLNGLQLRDAILKXNEDGTAY	1140
Db	1081	SGYQAKNTFISGDKWYFDDNNGYVMTGAQSTNGVNYFSLNGLQLRDAILKXNEDGTAY	1140
Qy	1141	YQNDGRRYENGYYQFMSGVWRHFNNGEMSVGLTVIDGVOYFDEMGOYQAKGFVTTADGK	1200
Db	1141	YQNDGRRYENGYYQFMSGVWRHFNNGEMSVGLTVIDGVOYFDEMGOYQAKGFVTTADGK	1200
Qy	1201	TRYFKQSGNMYRNFPIENEGKWLILGEDGAATVTSQTINGOHLVFRANGVOVKGFEVVT	1260
Db	1201	TRYFKQSGNMYRNFPIENEGKWLILGEDGAATVTSQTINGOHLVFRANGVOVKGFEVVT	1260
Qy	1261	DHHRISYDNGSGDOIIRNRFVRNAQGWYFDDNNGYAVTGARTINGOHLVFRANGVOVK	1320
Db	1261	DHHRISYDNGSGDOIIRNRFVRNAQGWYFDDNNGYAVTGARTINGOHLVFRANGVOVK	1320
Qy	1321	GBFVTDYGRISYDNGSGDOIIRNRFVRNAQGWYFDDNNGYAVTGARTINGOHLVFRAN	1380
Db	1321	GBFVTDYGRISYDNGSGDOIIRNRFVRNAQGWYFDDNNGYAVTGARTINGOHLVFRAN	1380
Qy	1381	GVQVKGFEVTDHHRISYDNGSGDOIIRNRFVRNAQGWYFDDNNGYAVTGARTINGOHL	1440
Db	1381	GVQVKGFEVTDHHRISYDNGSGDOIIRNRFVRNAQGWYFDDNNGYAVTGARTINGOHL	1440
Qy	1441	YFRANGVOVKGFEVTDYGRISYDNGSGDOIIRNRFVRNAQGWYFDDNNGYAVTGARTINGOHL	1475
Db	1441	YFRANGVOVKGFEVTDYGRISYDNGSGDOIIRNRFVRNAQGWYFDDNNGYAVTGARTINGOHL	1475

RESULT 15

AAU79284	AAU79284 standard; protein; 1476 AA.
XX	AAU79284;
AC	AAU79284;
DT	13-AUG-2002 (first entry)
XX	Streptococcus mutans monoclonal antibody-related protein #1.
DE	Antibody; dental caries; water insoluble glucan synthetase; anti-caries;
KW	glucosyl transferase-B; immunotherapy.
XX	Streptococcus mutans.
OS	Streptococcus mutans.
PN	JP2002114709-A.
XX	16-APR-2002.
PD	16-APR-2002.
XX	04-OCT-2000; 2000JP-00304889.
PF	04-OCT-2000; 2000JP-00304889.
XX	04-OCT-2000; 2000JP-00304889.
PR	04-OCT-2000; 2000JP-00304889.
XX	(UYNI-) UNIV NIPPON.
PA	(UYNI-) UNIV NIPPON.
XX	WPI; 2002-448885/48.
DR	WPI; 2002-448885/48.
XX	Anti-carries agent composed of a monoclonal antibody against an inhibitory
PT	enzyme against water insoluble glucan synthetase of glucosyl transferase-
PT	B (GTF-B) of Streptococcus mutans.
XX	Claim 3; Page 13-16; 28pp; Japanese.
PS	Claim 3; Page 13-16; 28pp; Japanese.
XX	The invention relates to a monoclonal antibody against dental caries and
CC	an anti-carries agent composed of a monoclonal antibody produced by
CC	

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 19:39:30 ; Search time 32.2941 Seconds
(without alignments)
3776.130 Million cell updates/sec

Title: US-10-797-821-34

Perfect score: 7741

Sequence: 1 MDKVRVYKLRVKRWTVS.....DRYGRISYYDANSRVRIN 1475

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/1aa/5 COMB.pep.*
- 2: /cgn2_6/ptodata/1/1aa/6 COMB.pep.*
- 3: /cgn2_6/ptodata/1/1aa/H COMB.pep.*
- 4: /cgn2_6/ptodata/1/1aa/PCTUS COMB.pep.*
- 5: /cgn2_6/ptodata/1/1aa/RE COMB.pep.*
- 6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7741	100.0	1475	2	US-09-007-999-2
2	7741	100.0	1475	2	US-09-210-361-2
3	7741	100.0	1475	2	US-09-740-274-2
4	5285	68.3	1375	2	US-09-210-361-4
5	5285	68.3	1375	2	US-09-740-274-4
6	3817.5	49.3	1430	2	US-09-008-172-2
7	3817.5	49.3	1430	2	US-09-210-361-6
8	3817.5	49.3	1430	2	US-09-740-274-6
9	3506	45.3	1577	1	US-08-793-824-2
10	2537	32.8	2057	2	US-09-499-203-2
11	2276	29.4	1278	2	US-09-604-957-3
12	2276	29.4	1278	2	US-09-995-749A-2
13	1685	21.8	545	2	US-09-604-957-4
14	1657	21.4	545	2	US-09-995-749A-10
15	1552	20.0	523	2	US-09-604-957-5
16	1547.5	20.0	522	2	US-09-995-749A-11
17	1379	17.8	535	2	US-09-604-957-7
18	1379	17.8	535	2	US-09-995-749A-13
19	1291.5	16.7	584	2	US-09-604-957-6
20	1289.5	16.7	584	2	US-09-995-749A-12
21	817	10.6	349	2	US-09-009-620-2
22	456	5.9	2710	1	US-08-480-604A-6
23	456	5.9	2710	1	US-08-405-496A-6
24	456	5.9	2710	2	US-08-915-136-6
25	456	5.9	2710	2	US-08-957-310-6
26	456	5.9	2710	2	US-10-011-366-6
27	456	5.9	2710	2	US-09-084-517-6

28	402	5.2	811	1	US-08-480-604A-7	Sequence 7, Appli
29	402	5.2	811	1	US-08-405-496A-7	Sequence 7, Appli
30	402	5.2	811	2	US-08-915-136-7	Sequence 7, Appli
31	402	5.2	811	2	US-08-957-310-7	Sequence 7, Appli
32	402	5.2	811	2	US-10-011-366-7	Sequence 7, Appli
33	402	5.2	811	2	US-09-084-517-7	Sequence 7, Appli
34	402	5.2	812	1	US-08-480-604A-29	Sequence 29, Appl
35	402	5.2	812	2	US-08-915-136-29	Sequence 29, Appl
36	402	5.2	812	2	US-09-084-517-29	Sequence 29, Appl
37	400	5.2	866	2	US-09-545-773-2	Sequence 2, Appli
38	400	5.2	866	2	US-10-222-038-2	Sequence 2, Appli
39	381	4.9	2366	1	US-08-480-604A-10	Sequence 10, Appl
40	381	4.9	2366	1	US-08-405-496A-10	Sequence 10, Appl
41	381	4.9	2366	2	US-08-915-136-10	Sequence 10, Appl
42	381	4.9	2366	2	US-08-957-310-10	Sequence 10, Appl
43	381	4.9	2366	2	US-10-011-366-10	Sequence 10, Appl
44	381	4.9	2366	2	US-09-084-517-10	Sequence 10, Appl
45	369.5	4.8	1231	2	US-08-714-741-41	Sequence 41, Appl

ALIGNMENTS

RESULT 1
US-09-007-999-2
; Sequence 2, Application US/09007999
; Patent No. 6087559
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0356D
; CURRENT APPLICATION NUMBER: US/09/007,999
; CURRENT FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-007-999-2

Query Match	100.0%	Score 7741;	DB 2;	Length 1475;
Best Local Similarity	100.0%	Pred. No. 0;	Mismatches 0;	Indels 0; Gaps 0;
Matches 1475;	Conservative	0;		
QY	1	MDKVRVYKLRVKRWTVSVASAVMTLTTLGGGLVKADSNESKQISNDSNTSVVTANE	60	
Db	1	MDKVRVYKLRVKRWTVSVASAVMTLTTLGGGLVKADSNESKQISNDSNTSVVTANE	60	
QY	61	ESNVITEATSKQEAASSQTNHTVTSSSTSVVNPKEVSVNPYTVGTASNGEKLQNTT	120	
Db	61	ESNVITEATSKQEAASSQTNHTVTSSSTSVVNPKEVSVNPYTVGTASNGEKLQNTT	120	
QY	121	TVDKTSEAAANNISKQTTTADTDVDDNNAANLQLEKLPNVKEIDGKYYVYDNNKQVPT	180	
Db	121	TVDKTSEAAANNISKQTTTADTDVDDNNAANLQLEKLPNVKEIDGKYYVYDNNKQVPT	180	
QY	181	NFTLIADGKILHFDGTGAYTDTSDTVNKDITVTRSNLYKKYNQVYDRAQSFVHVDHYL	240	
Db	181	NFTLIADGKILHFDGTGAYTDTSDTVNKDITVTRSNLYKKYNQVYDRAQSFVHVDHYL	240	
QY	241	TAESWYRPKYILKDGKVTWTQSTEKDFRPLMTWPDQETQRYVNMNAQLGINKTYDDT	300	
Db	241	TAESWYRPKYILKDGKVTWTQSTEKDFRPLMTWPDQETQRYVNMNAQLGINKTYDDT	300	
QY	301	SNQLQNLNATAATQIAKIEAKITTLKNTDHLQRTISAFVKTSQAWNSDSEKPPDDHLQNGA	360	
Db	301	SNQLQNLNATAATQIAKIEAKITTLKNTDHLQRTISAFVKTSQAWNSDSEKPPDDHLQNGA	360	
QY	361	VLYDNEGKLTTPVANSNYRILNRTPTNQTKKDPRTADTWTIGGYEFLLANDVDSNPVQV	420	

Db 361 VLYDNEGKLTTPYANSNYRILNTPNTQCKDPRYTADNTIGGYEFLANDVDNSNPVQ 420
QY 421 AEQLNLHFLMNFNGNIYANDPDANFDSIRVDAVDNVDADLLQIAGDYLKAAGIHKNDKA 480
Db 421 AEQLNLHFLMNFNGNIYANDPDANFDSIRVDAVDNVDADLLQIAGDYLKAAGIHKNDKA 480
QY 481 ANDHLSILEAWSNDTPYLDHDDGNMNDKRLSLLSLAKPLNQRSGMNPPLITNSLV 540
Db 481 ANDHLSILEAWSNDTPYLDHDDGNMNDKRLSLLSLAKPLNQRSGMNPPLITNSLV 540
QY 541 NRTDNaETAAPVPSYFIRAHDSVQDLDIADIIKAEINPNVGVSYFTMEIEKKAPEIYNK 600
Db 541 NRTDNaETAAPVPSYFIRAHDSVQDLDIADIIKAEINPNVGVSYFTMEIEKKAPEIYNK 600
QY 601 DLLATEKKYTHYNTALSALLTNKSSPRVYVYGMFTDDGOYMAHKTINYEAIETLLKA 660
Db 601 DLLATEKKYTHYNTALSALLTNKSSPRVYVYGMFTDDGOYMAHKTINYEAIETLLKA 660
QY 661 RIKYVSGGQAMRNQOVGNSEIITSVRYGKGAALKATDGTDRTRTSGVAVIEGNPRLK 720
Db 661 RIKYVSGGQAMRNQOVGNSEIITSVRYGKGAALKATDGTDRTRTSGVAVIEGNPRLK 720
QY 721 ASDRVVNVNGAAHKNQAYRPLLLTTDNGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKG 780
Db 721 ASDRVVNVNGAAHKNQAYRPLLLTTDNGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKG 780
QY 781 YANPOVSGYLVGVVPGGAALIKWFALRLARPHQOMASVHQNAALDSRVNMFEGFSNFQAPA 840
Db 781 YANPOVSGYLVGVVPGGAALIKWFALRLARPHQOMASVHQNAALDSRVNMFEGFSNFQAPA 840
QY 841 TKKEEYTNVIAKNVDKFAEKGVTDFEMAPQVVSSTDSGFLDSVTQNGYAFTRDRVDLGIS 900
Db 841 TKKEEYTNVIAKNVDKFAEKGVTDFEMAPQVVSSTDSGFLDSVTQNGYAFTRDRVDLGIS 900
QY 901 KPNKGTADDLVKAIALKHSKGIKVMADVPDQMYAFPEKEVVTATRVKDTPTVAGSQI 960
Db 901 KPNKGTADDLVKAIALKHSKGIKVMADVPDQMYAFPEKEVVTATRVKDTPTVAGSQI 960
QY 961 KNTLVVVDGSKGDKQQAQYKGAFLLEELQAKYPFELFARKQISTGVMPDPSVKIKOWSAKY 1020
Db 961 KNTLVVVDGSKGDKQQAQYKGAFLLEELQAKYPFELFARKQISTGVMPDPSVKIKOWSAKY 1020
QY 1021 FNGTNILGRGAGYVLKDQATNTYFNI SDNKEINFLPKTLNODSQVGSYDGKGVYVYST 1080
Db 1021 FNGTNILGRGAGYVLKDQATNTYFNI SDNKEINFLPKTLNODSQVGSYDGKGVYVYST 1080
QY 1081 SGYQAKNTFISGDKWYFDNNGYVMTGAQSLNGVNYVFLSNGQLRDAILLKNEGDGTAY 1140
Db 1081 SGYQAKNTFISGDKWYFDNNGYVMTGAQSLNGVNYVFLSNGQLRDAILLKNEGDGTAY 1140
QY 1141 YGNDGRRYENGYYQPMGSVWRHFNNGEMSVGLTVIDGQVQYFDEMGYQAKGFVTTADGK 1200
Db 1141 YGNDGRRYENGYYQPMGSVWRHFNNGEMSVGLTVIDGQVQYFDEMGYQAKGFVTTADGK 1200
QY 1201 IRYFDKQSGMYRNRPIENEBGKWLVLGEDGAATVGSOTINGQHLYFRANGVQVKGFEVFT 1260
Db 1201 IRYFDKQSGMYRNRPIENEBGKWLVLGEDGAATVGSOTINGQHLYFRANGVQVKGFEVFT 1260
QY 1261 DHGGRISYYDNGSGDOI RNRFRVNAQOGWFYFDNNGYAVTGARTINGQLLYFRANGVQVK 1320
Db 1261 DHGGRISYYDNGSGDOI RNRFRVNAQOGWFYFDNNGYAVTGARTINGQLLYFRANGVQVK 1320
QY 1321 GEFVTRDGRISYYDNGSGDOI RNRFRVNAQOGWFYFDNNGYAVTGARTINGQHLYFRAN 1380
Db 1321 GEFVTRDGRISYYDNGSGDOI RNRFRVNAQOGWFYFDNNGYAVTGARTINGQHLYFRAN 1380
QY 1381 GVQVKGFEVTRDGRISYYDNGSGDOI RNRFRVNAQOGWFYFDNNGYAVTGARTINGQHL 1440
Db 1381 GVQVKGFEVTRDGRISYYDNGSGDOI RNRFRVNAQOGWFYFDNNGYAVTGARTINGQHL 1440
QY 1441 YFRANGVQVKGFEVTRDGRISYYDANSGERVRIN 1475
Db 1441 YFRANGVQVKGFEVTRDGRISYYDANSGERVRIN 1475

Db 1441 YFRANGVQVKGFEVTRDGRISYYDANSGERVRIN 1475
RESULT 2
US-09-210-361-2
; Sequence 2, Application US/09210361
; Patent No.: 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-210-361-2
Query Match 100.0%; Score 7741; DB 2; Length 1475;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1475; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MDKVVYKLRVKRWRVTVSASAVMTLTTLGGVLKADSNESKQISDNTSNTSVTANE 60
Db 1 MDKVVYKLRVKRWRVTVSASAVMTLTTLGGVLKADSNESKQISDNTSNTSVTANE 60
QY 61 ESNVITEATSKQEAASSQTNHTVTSSSTSVNPKVSNPVTYVGTASNGEKLQNTT 120
Db 61 ESNVITEATSKQEAASSQTNHTVTSSSTSVNPKVSNPVTYVGTASNGEKLQNTT 120
QY 121 TVDKTSEAAANNISKQTEADTDVIDDSNAANLQILEKLPNVKEIDGKYVYDNNKGVRT 180
Db 121 TVDKTSEAAANNISKQTEADTDVIDDSNAANLQILEKLPNVKEIDGKYVYDNNKGVRT 180
QY 181 NFTLIADGKILHFDGTAYTDTSIDTVNKDITVTRSNLYKKYQVYDRSAQSFEHVDHYL 240
Db 181 NFTLIADGKILHFDGTAYTDTSIDTVNKDITVTRSNLYKKYQVYDRSAQSFEHVDHYL 240
QY 241 TAESWYRKYILKDGKWTQSTQTEKDFRLMTWMPDQETORQYVNMNAQLGINKTYDDT 300
Db 241 TAESWYRKYILKDGKWTQSTQTEKDFRLMTWMPDQETORQYVNMNAQLGINKTYDDT 300
QY 301 SNQLQNTAAATIOAKIEAKITTLKNTDWTQRTISAFVKTSAMNSDSEKPPDDHLONGA 360
Db 301 SNQLQNTAAATIOAKIEAKITTLKNTDWTQRTISAFVKTSAMNSDSEKPPDDHLONGA 360
QY 361 VLYDNEGKLTTPYANSNYRILNRTPTNQTGKKDPRYTADNTIGGYEFLANDVDNSNPVQ 420
Db 361 VLYDNEGKLTTPYANSNYRILNRTPTNQTGKKDPRYTADNTIGGYEFLANDVDNSNPVQ 420
QY 421 AEQLNLHFLMNFNGNIYANDPDANFDSIRVDAVDNVDADLLQIAGDYLKAAGIHKNDKA 480
Db 421 AEQLNLHFLMNFNGNIYANDPDANFDSIRVDAVDNVDADLLQIAGDYLKAAGIHKNDKA 480
QY 481 ANDHLSILEAWSNDTPYLDHDDGNMNDKRLSLLSLAKPLNQRSGMNPPLITNSLV 540
Db 481 ANDHLSILEAWSNDTPYLDHDDGNMNDKRLSLLSLAKPLNQRSGMNPPLITNSLV 540

Qy	541	NRTDDNAETAAVPSYSPFIRAHDSQVODLIADI	IIKAEINPNVVGVSFTMBEIKKQAFIYNK	600
Db	541	NRTDDNAETAAVPSYSPFIRAHDSQVODLIADI	IIKAEINPNVVGVSFTMBEIKKQAFIYNK	600
Qy	601	DLLEATEKKYTHYNTALSYALLLTKNKSSVPRVYVYGD	MTDDCGQYMAHKNTINYEATLTKKA	660
Db	601	DLLEATEKKYTHYNTALSYALLLTKNKSSVPRVYVYGD	MTDDCGQYMAHKNTINYEATLTKKA	660
Qy	661	RIKYVSGGQAMRNQOVGNSEIITSVRYGKGALKATD	TGDRTRTSGVAVIEGNPNPSRLK	720
Db	661	RIKYVSGGQAMRNQOVGNSEIITSVRYGKGALKATD	TGDRTRTSGVAVIEGNPNPSRLK	720
Qy	721	ASDRVVVMGAAHKKNQAYRPLLLTTDNGIKAYHSD	QEAAGLVRYTNDRGELIFTAADIKG	780
Db	721	ASDRVVVMGAAHKKNQAYRPLLLTTDNGIKAYHSD	QEAAGLVRYTNDRGELIFTAADIKG	780
Qy	781	YANPOVSGYLGVMYPVPGAALIKMFALRLARPHQOM	ASVHQNAALDSRVMEGFSNFQAF	840
Db	781	YANPOVSGYLGVMYPVPGAALIKMFALRLARPHQOM	ASVHQNAALDSRVMEGFSNFQAF	840
Qy	841	TKKSEYTNVJIAKNVDKFAEKGVTDFEMAPQVSG	STDGSLPDSVIQNGYAFTRDYDLGIS	900
Db	841	TKKSEYTNVJIAKNVDKFAEKGVTDFEMAPQVSG	STDGSLPDSVIQNGYAFTRDYDLGIS	900
Qy	901	KPNKYGTADDLVKAIKALHSGKIKVMADWPQDMYAF	PEKEVVTATRVDKYGTVPVAGSQI	960
Db	901	KPNKYGTADDLVKAIKALHSGKIKVMADWPQDMYAF	PEKEVVTATRVDKYGTVPVAGSQI	960
Qy	961	KNTLYVVDGSKSGKQQAQYGGAFLEELQAKYPEL	FARKQISTQVPMDPSPVKIKQWSAKY	1020
Db	961	KNTLYVVDGSKSGKQQAQYGGAFLEELQAKYPEL	FARKQISTQVPMDPSPVKIKQWSAKY	1020
Qy	1021	PNGTNIILGRGAGYVLKQOATNTYFNI	SDNKEINF.LP.KTLLINODSQVGSVDGKYVYVYST	1080
Db	1021	PNGTNIILGRGAGYVLKQOATNTYFNI	SDNKEINF.LP.KTLLINODSQVGSVDGKYVYVYST	1080
Qy	1081	SGYQAKNTFISEGDKWYVYFDNNGYMWVTGAOS	INGWNYVYF.LSNG.LQLRDALLKNEDGTYAY	1140
Db	1081	SGYQAKNTFISEGDKWYVYFDNNGYMWVTGAOS	INGWNYVYF.LSNG.LQLRDALLKNEDGTYAY	1140
Qy	1141	YGNDRRYENGYOYFMGSGVWRHFNNGEMSVGLT	VIDGQVQYFDEMGYQAKGKFVTTADGK	1200
Db	1141	YGNDRRYENGYOYFMGSGVWRHFNNGEMSVGLT	VIDGQVQYFDEMGYQAKGKFVTTADGK	1200
Qy	1201	IRYPDKOSGNMYRNPFTIENBEGKWLYGEGDGA	AVTGSQTTINGQHL.YFRANGVQVKGFBVT	1260
Db	1201	IRYPDKOSGNMYRNPFTIENBEGKWLYGEGDGA	AVTGSQTTINGQHL.YFRANGVQVKGFBVT	1260
Qy	1261	DHGHGRIISYDGNNSGDOI.RNRFVRNAQOGWF	YFDNNGYAVTGARTINGOHL.YFRANGVQVK	1320
Db	1261	DHGHGRIISYDGNNSGDOI.RNRFVRNAQOGWF	YFDNNGYAVTGARTINGOHL.YFRANGVQVK	1320
Qy	1321	GEFVTDYGRISYDGNNSGDOI.RNRFVRNAQOG	WFYFDNNGYAVTGARTINGOHL.YFRAN	1380
Db	1321	GEFVTDYGRISYDGNNSGDOI.RNRFVRNAQOG	WFYFDNNGYAVTGARTINGOHL.YFRAN	1380
Qy	1381	GVQVKGFBVTDYDRHGRISYDGNNSGDOI.RN	RFVRNAQOGWFYFDNNGYAVTGARTINGOHL	1440
Db	1381	GVQVKGFBVTDYDRHGRISYDGNNSGDOI.RN	RFVRNAQOGWFYFDNNGYAVTGARTINGOHL	1440
Qy	1441	YFRANGVQVKGFBVTDYGRISYDGNNSGGRV	IN 1475	
Db	1441	YFRANGVQVKGFBVTDYGRISYDGNNSGGRV	IN 1475	

RESULT 3

US-09-740-274-2

; Sequence 2, Application US/09740274

; Patent No. 6465203

GENERAL INFORMATION:

APPLICANT: Nichpls, Scott E.

TITLE OF INVENTION: Glucan-containing Compositions and Paper

Qy	661	RIKTVSGQAMRNQOVGNSEIITSVRVYKGALKATDGTGDRTRTFTSGVAVIEGNNPSLRK	720
Db	661	RIKTVSGQAMRNQOVGNSEIITSVRVYKGALKATDGTGDRTRTFTSGVAVIEGNNPSLRK	720
Qy	721	ASDRVVNMGAHKNQAYRPLLLTTDNGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKG	780
Db	721	ASDRVVNMGAHKNQAYRPLLLTTDNGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKG	780
Qy	781	YANPQVSGYLGVWVPVGAALIKMFALRLARPHQOMASVHQNAALDSRVMEGFSNFOAFA	840
Db	781	YANPQVSGYLGVWVPVGAALIKMFALRLARPHQOMASVHQNAALDSRVMEGFSNFOAFA	840
Qy	841	TKKBEYTNVVIKXNDVFAEWGVTDFEMAPOYVSGTSGFSLDSVIQNGYAFTRDYDLGIS	900
Db	841	TKKBEYTNVVIKXNDVFAEWGVTDFEMAPOYVSGTSGFSLDSVIQNGYAFTRDYDLGIS	900
Qy	901	KPNKYGTADDLVKAIKALHSGIKVMADWPDQWYAFPEKEVVTATRVDKYGPVPAGSQI	960
Db	901	KPNKYGTADDLVKAIKALHSGIKVMADWPDQWYAFPEKEVVTATRVDKYGPVPAGSQI	960
Qy	961	KNTLYVVVDGKSSGKQOQAKYGGAFLBEOAQYPELFARKQISTGVPMDPSPVKIKOWSAKY	1020
Db	961	KNTLYVVVDGKSSGKQOQAKYGGAFLBEOAQYPELFARKQISTGVPMDPSPVKIKOWSAKY	1020
Qy	1021	PNGTNIILGRGAGYVLDQAANTYFNISDNKEINFLPKTLLNQDSQVGFSDGKGYYVYST	1080
Db	1021	PNGTNIILGRGAGYVLDQAANTYFNISDNKEINFLPKTLLNQDSQVGFSDGKGYYVYST	1080
Qy	1081	SGYQAKNTFISEGDKWYTFDNNGYMVTGAOSINGVNYFSLNGIQLRDAILKKNEDGTIYAY	1140
Db	1081	SGYQAKNTFISEGDKWYTFDNNGYMVTGAOSINGVNYFSLNGIQLRDAILKKNEDGTIYAY	1140
Qy	1141	YGNDGRYENGYYOFMSGVWRHFNNGEMSVGLTVIDGQVOYFDEMGYOAKGKFVTADGK	1200
Db	1141	YGNDGRYENGYYOFMSGVWRHFNNGEMSVGLTVIDGQVOYFDEMGYOAKGKFVTADGK	1200
Qy	1201	IRYPDKOSGNNYRNFTENEGKWLVLGEQGAATVGSQTTINGOHLVFRANGVOVKGFEVFT	1260
Db	1201	IRYPDKOSGNNYRNFTENEGKWLVLGEQGAATVGSQTTINGOHLVFRANGVOVKGFEVFT	1260
Qy	1261	DHGRISYYDGNSGDQIRNRFVRNAQOWFYFDNNGYAVTGARTINGQLLYFRANGVOVK	1320
Db	1261	DHGRISYYDGNSGDQIRNRFVRNAQOWFYFDNNGYAVTGARTINGQLLYFRANGVOVK	1320
Qy	1321	GEFVTDYRGRISYYDGNSGDQIRNRFVRNAQOWFYFDNNGYAVTGARTINGOHLVFRAN	1380
Db	1321	GEFVTDYRGRISYYDGNSGDQIRNRFVRNAQOWFYFDNNGYAVTGARTINGOHLVFRAN	1380
Qy	1381	GVQVKGFEVTDHGRISYYDGNSGDQIRNRFVRNAQOWFYFDNNGYAVTGARTINGOHL	1440
Db	1381	GVQVKGFEVTDHGRISYYDGNSGDQIRNRFVRNAQOWFYFDNNGYAVTGARTINGOHL	1440
Qy	1441	YFRANGVOVKGEFVTDYRGRISYYDANSGERVRIN	1475
Db	1441	YFRANGVOVKGEFVTDYRGRISYYDANSGERVRIN	1475

RESULT 4

US-09-210-361-4
 ; Sequence 4, Application US/09210361
 ; Patent No. 6284479
 ; GENERAL INFORMATION:
 ; APPLICANT: Nichols, Scott E.
 ; TITLE OF INVENTION: Substitutes for Modified Starches and
 ; TITLE OF INVENTION: Latexes in Paper Manufacture
 ; FILE REFERENCE: 0357CR
 ; CURRENT APPLICATION NUMBER: US/09/210,361
 ; CURRENT FILING DATE: 1998-12-11
 ; EARLIER APPLICATION NUMBER: 09/007,999
 ; EARLIER FILING DATE: 1998-01-16
 ; EARLIER APPLICATION NUMBER: 08/478,704
 ; EARLIER FILING DATE: 1995-06-07
 ; EARLIER APPLICATION NUMBER: 09/009,620

Db 775 IKAYHSDQBAAGLVRVTNRDRELIFTAADIKGYANPQVSGYLGWVPVGAADQ--DVRV 832
Qy 809 A---RPHQQWASVHQAALDSRVMPFGFSNFQAFATKCEYTNVIAKNVDKFAEWGVT 865
Db 833 AASTARSTGKSHQAALDSRVMPFGFSNFQAFATKCEYTNVIAKNVDKFAEWGVT 892
Qy 866 FEMAPQVYSTGDSFLDSVIQNGYATDRYDLGISKPNKYGTADDLVKAIKALHSGIKV 925
Db 893 FEMAPQVYSTGDSFLDSVIQNGYATDRYDLGISKPNKYGTADDLVKAIKALHSGIKV 952
Qy 926 MADWVPDQMYAPPEKEVVTATRVKYGTPVAGSIOIKNTLYVVDGKSSGKQQAQKYGAE 985
Db 953 MADWVPDQMYALPEKEVVTATRVKYGTPVAGSIOIKNTLYVVDGKSSGKQQAQKYGAE 1012
Qy 986 EELQAKYPELFARKQISTGVPMDPSVKIKQWSAKYFNGNIIILGRGAGYVLKQATNTYFN 1045
Db 1013 EELQAKYPELFARKQISTGVPMDPSVKIKQWSAKYFNGNIIILGRGAGYVLKQATNTYFN 1072
Qy 1046 -ISDNKEINFLPKTLN-----QDSQVGSYDGKGVVYSTSGYQAKNTFISEGDKWYF 1099
Db 1073 LVSDN---TFLPKSLVNPNGHTSSVTGLVFDGKGVVYSTSGQAKNAFISLGNWYF 1129
Qy 1100 DNNQYMTCAQSTNGWNYFLSGLQRLDAILKNDGTAYVYCNDRRYENGYQPMGV 1159
Db 1130 DNNQYMTCAQSTNGWNYFLSGLQRLDAILKNDGTAYVYCNDRRYENGYQPMGV 1188
Qy 1160 WRHFNNGEMSVGLTIDGQVQFDEMGYQAKGFVTADCKIRYFDKQSGNRYNRFIEN 1219
Db 1189 WRYFQNGINAVGLTRVHGAVQVFDASGFAKQGFITADGKLYFRDQSGNQSIRFVRN 1248
Qy 1220 EBGKMLYLGEAAVTGQSTINGOHLFRANGVQVKGFEVTPDHGHSIYDNGSGDQIRN 1279
Db 1249 SKGEWFLDHNGVAVTGTVTENGRLYFKPNGVQAKGEFIRDANGLYLYDPSNGNEVEN 1308
Qy 1280 RFRNAQCGWFYEDNNGYAVTGARTING-----QLLYPR 1313
Db 1309 RFRNRSKGWFLFDHNGIAVTGARVVVNGHASILSLMVFR 1347

RESULT 5
US-09-740-274-4
; Sequence 4, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-4

Query Match 68.3%; Score 5285; DB 2; Length 1375;
Best Local Similarity 75.9%; P-red. No. 0;

Matches 1031; Conservative 101; Mismatches 169; Indels 58; Gaps 14;
Qy 1 MDKVRKYLKRVKRWVTVSVASAVMTLTLTSLGVLKADSNESKSIQSNDSNTSVVTANE 60
Db 1 MEKKVRKFLKRVKRWVTVSVASAVMTLTLTSLGVLKADSNESKSIQSNDSNTSVVTANE 60
Qy 61 -----ESNVITEATSKQBAASQTNHTVTSSSTSVVNPKEWENPYTVGSETA-- 109
Db 61 AAKETLTATDTSTATSATSQPTA--TVTDNVSTTNGSTNTTANTANFVVKPTTTSQAKT 118
Qy 110 SNGEKLQNTTVDKTSSEA-----AANNISKOTTEAD-----TDVDDSN 149
Db 119 DNSDKIITTSKAVNRLTATGKFPANNTAHPKTVTDKIPIKPKIKGLKQPSLSQDDI 178
Qy 150 AANLQILEKLPVKIEDGKYVYDNNKVRTMTFLIADGKILHFDSTGATDTSIDTVNK 209
Db 179 AA----LGNVKIRKNGKYYKEDGTLQKYNALNINIKTFFBFTGALSNNLPSKKG 234
Qy 210 DIVTT-RGNLYKKYQVYDRSQAQSFHVDHYLTAEVWYRPKYILKDGKTWTOSTEKDPRP 268
Db 235 NITNDNTNSPAQYNQVSTVDVANFEHVDHYLTAEVWYRPKYILKDGKTWTOSTEKDPRP 294
Qy 269 LLMTWPDQETORQVYVNMNAOLGINKYDDTSNOIOLNIAAATQAKIEAKITTLKNTD 328
Db 295 LLMTWPDQETORQVYVNMNAOLGINKYDDTSNOIOLNIAAATQAKIEAKITTLKNTD 354
Qy 329 WLQRTISAFVKTQSASWNSDEKPPDDHLQNGAVLYDNEGKLTPTYANSNRYILNRTPTNQT 388
Db 355 WLQRTISAFVKTQSASWNSDEKPPDDHLQNGAVLYDNEGKLTPTYANSNRYILNRTPTNQT 414
Qy 389 GKDKPRYTADNTIGGYEFLLANDVNSNPVQAEOLNMLHFLMNFNIGNIYANDPDANFDSI 448
Db 415 GKDKPRYTADNTIGGYEFLLANDVNSNPVQAEOLNMLHFLMNFNIGNIYANDPDANFDSI 474
Qy 449 RVDADVNDADLLQIAGDYLKAAGIHKNDKAANDHLSILEAWSNDTPYLHDDGDNMTN 508
Db 475 RVDADVNDADLLQIAGDYLKAAGIHKNDKAANDHLSILEAWSNDTPYLHDDGDNMTN 534
Qy 509 MDNKLRLSLFLSLAKPLNQRSGMNPILTNSLVNRTDDNAETAAPVPSYFIRAHDSVQDL 568
Db 535 MDNKLRLSLFLSLAKPLNQRSGMNPILTNSLVNRTDDNAETAAPVPSYFIRAHDSVQDL 594
Qy 569 IADIIKAEINPNVVGYSFTMBEIKKAFIYNKDLATEKKYTHYNTALSYALLNTKSSV 628
Db 595 IRNIIRTEINPNVVGYSFTMBEIKKAFIYNKDLATEKKYTHYNTALSYALLNTKSSV 654
Qy 629 PRVYVGMFTDDGQYMAHKTINYEAEITLLKARIKYVSGQAMRNQOVGNSBIITSVRYG 688
Db 655 PRVYVGMFTDDGQYMAHKTINYEAEITLLKARIKYVSGQAMRNQOVGNSBIITSVRYG 714
Qy 689 KGALKATDGTDRTRTSGVAVTEGNPSILKASDRVVVNMAAHKQAYRPLLLTTDNG 748
Db 715 KGALKATDGTDRTRTSGVAVTEGNPSILKASDRVVVNMAAHKQAYRPLLLTTDNG 774
Qy 749 IKAYHSDQBAAGLVRVTNRDRELIFTAADIKGYANPQVSGYLGWVPVGAALIKMFA 808
Db 775 IKAYHSDQBAAGLVRVTNRDRELIFTAADIKGYANPQVSGYLGWVPVGAALIKMFA 832
Qy 809 A---RPHQQWASVHQAALDSRVMPFGFSNFQAFATKCEYTNVIAKNVDKFAEWGVT 865
Db 833 AASTARSTGKSHQAALDSRVMPFGFSNFQAFATKCEYTNVIAKNVDKFAEWGVT 892
Qy 866 FEMAPQVYSTGDSFLDSVIQNGYATDRYDLGISKPNKYGTADDLVKAIKALHSGIKV 925
Db 893 FEMAPQVYSTGDSFLDSVIQNGYATDRYDLGISKPNKYGTADDLVKAIKALHSGIKV 952
Qy 926 MADWVPDQMYAPPEKEVVTATRVKYGTPVAGSIOIKNTLYVVDGKSSGKQQAQKYGAE 985
Db 953 MADWVPDQMYALPEKEVVTATRVKYGTPVAGSIOIKNTLYVVDGKSSGKQQAQKYGAE 1012
Qy 986 EELQAKYPELFARKQISTGVPMDPSVKIKQWSAKYFNGNIIILGRGAGYVLKQATNTYFN 1045
Db 1013 EELQAKYPELFARKQISTGVPMDPSVKIKQWSAKYFNGNIIILGRGAGYVLKQATNTYFN 1072


```
; FILING DATE: 24-AUG-1994
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 61 2 9957 5944
; TELEFAX: 61 2 957 6288
; TELEX: 26547
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1577 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus salivarius
; US-08-793-824-2

Query Match
Best Local Similarity 45.3%; Score 3506; DB 1; Length 1577;
Matches 733; Conservative 239; Mismatches 468; Indels 162; Gaps 28;

QY 1 MDKKVRYKLRKYKRWTVSVASAVMTLFTLSGLV-----KADSNEKS-----QISN 49
DB 1 MENKVRFKLHKYKONWTVIGVTT--LSMVALAGGSLLAQGVKVEADETSAPNGDGLQQLSE 58

QY 50 DSNTSVVTANESNVITEATSKQ-----EASQTNHTVTTSSSTSV 92
DB 59 DGTASLVT-----TTVTQASQAQASVAVATASVSHTSFQAATSVAVSQEAQAQTSVP 114

QY 93 VNPKEVNSNPYTVGTASNGEKLQ-----116
DB 115 ASQEVAVSS-----QTQSSGQETQTEQVSQQTSTQVAGQTSQSTPSVTSQARPRVLT 169

QY 117 -----NOTTVDKTSAAANNI-----SKOTTEAD 141
DB 170 NAAPAIAATRAADSTIRANRWNTITASGTTPTNVTIITGNTPKPNVTVTSPNGTRPN 229

QY 142 TDVIDDSNAANLQILEKL P-----NVKEIDGKYVYDNNKGVRT 180
DB 230 VTIVTQPNQPNKPVQSPQSPQPNQPNQSLDKPVASNLKLTIDGKQYVE-NGVVKK 288

QY 181 NFTLADGKILHFDGTGATDTSIDTVNKDITVTTSNLYKKYQNVYDRSAQSFEHVDHYL 240
DB 289 NAAIELDGLRYFDGTGAWVQSKPLRYADAIIPNNS-IYAVYNQAYDTSKSSFEHLDNFL 347

QY 241 TAEWYRPKYILKDGKTWTSQTEKDFRPLMTWPDQETQRYQVNYMAQLGINKTYDDT 300
DB 348 TADSWYRPKQILKDGKNWTASTEKDYRPLMTWPDQETQRYQVNYMAQLGINKTYDDT 407

QY 301 SNQLQNLTAATIAQIEAKITTLKNTDWRQTISAFVKTSAMNSDSEKPP---DDHLQ 357
DB 408 MMSYDLAAAAETVQRIEERIGREGNTTTLRLQMSDFIKTQPGWNSSEEDNLLVGKDLHQ 467

QY 358 NGAVLYDNKGKLTYPANSNYRLARTPTNOTKQKDPRTADNTICGYEFLANDVDNSNP 417
DB 468 GGAULTFLN-NSATSHANSDFRLMNRTPNTQGTGR--KYHIDRSNGGYELLANDIDNSNP 524

QY 418 VVQAEQLNWLHFLMNFNIYANDPANFDSIRVDADVNDVADLLQIAGDYLKAAKGIHKN 477
DB 525 AVQAEQLNWLHYIMNIGILNDPNSANFDGVRIDAVNDVADLLQIASDYFKEKTRVADN 584

QY 478 DKAANDHLSILEAMSDNTPYLHDDGDNMIMNDKRLSLLSFLSAKPLNQRSGMNPLITN 537
DB 585 EANAIAHLSILEAWSYNDHQNKQPKGAQLSIDNPLRETLTTLFLRKSNGYRSLERVITN 644

QY 538 SLVNRTDNAETAAPVPSFSFRAHDSVODLIADIKAIEINPVNVGVYFTWEEIKKAFEI 597
DB 645 SLNRRSSQKHPTPDANFIYFRAHDSVQAVLANIISKQINPKTDGFTTMDLQKAFEI 704

QY 598 YNKDLATEKYYTHYNTALSVALLLTNKSSVPRVYVYDGMFTDDGQYMAHKINTYEAETL 657
DB 705 YNADIADKADKYQTQNIIPAAATMLTNKDSITRVYVYDGLFTDDGQYMAEKSPYNAIDAL 764

QY 658 LKARIKYVSGQAMRNQGVNSEIITSVRYGKALKATDGTGTRTTRTSQVAVIEGNPNSL 717
```

RESULT 10

```
US-09-499-203-2
; Sequence 2, Application US/09499203
; Patent No. 6570065
; GENERAL INFORMATION:
; APPLICANT: KOSSMANN, Jens
; APPLICANT: WELSH, Thomas
; APPLICANT: QUANZ, Martin
; APPLICANT: KNUTH, Karola
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternansucrase
; FILE REFERENCE: 147-196P
; CURRENT APPLICATION NUMBER: US/09/499,203
; CURRENT FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2057
```

; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-09-499-203-2

Query Match 32.8%; Score 2537; DB 2; Length 2057;
Best Local Similarity 39.8%; Pred. No. 2.4e-161;
Matches 607; Conservative 205; Mismatches 486; Indels 228; Gaps 46;

QY 93 VNPKEVSNPVYTGTSANGE----KLQNTTTVDKTSSEAANNISKOTTREADTVIDDS 148
DB 286 VNNKRIYFAPNT-GNAVANTEIINGKLQGRDANGNQVKNAPSKDVAGNTFYFDANGV--- 341
QY 149 NAANLQILLEKLPVKEIDKYYVYDNNNGKVRTNFTLIADGKLLHPD-ETGAYTDSIDTV 207
DB 342 -----MLTGLTISGKTYLDEQGHLRKNYAGTFFNQPMFYDADTGA-GKTAIEYQ 391
QY 208 NKDITVTRSNLYKKN--QVYDRSAQSPEHVDHYLTAEBSWYRKPVKILKDGKTKWTOSTEKD 265
DB 392 PDQGLVSQSNENTPHNAASKSYDKS--SPENVGGLYLTADTWYRPTDILKNGDWTASTETD 449
QY 266 FRPLMTWMPQDETQORQVYVYNNAO-LGINKTYDDTSNQLQNLIAAAATIQAKIEAKITTL 324
DB 450 MRPLMTWMPDQKTOQANLNFMSKGLGITTYYTAATSKTLNDAAFVIQTAIEQOISLK 509
QY 325 KNTDWLQRTISAFVKTQSAWNSDSE-KPFD--DHLONGAVLYDNEGKLTTPYANS-NYRIL 380
DB 510 KSTEWLRDAIDSFVKTOANWNRKQTEDEAFGLQWLQGGFLAYQDDSHRTPNPTDGNRRKL 569
QY 381 NRTPNQTKKDPRTYADNTTIGGYEFFLLANDVDSNPVVQAEQLNWLHFLMFGNIYAND 440
DB 570 GROPINIGSKD---TTGK--GSEFLLANDIDNSNPVQAEQLNWLHFLMFGSITGN 624
QY 441 PDANFDSIRDAVDNVDADLLQIAGDYLLKAAKGTHKNDKAANDHLSLEAMSDDNDTPYLH 500
DB 625 DNANFDGIRDAVDNVDADLLKIAAGDYFKALYGTDKSDANANKHLSILEDDWNGKDPQVYN 684
QY 501 DDGDMNIMNDKRLSLFLSLAKPLNQRSGM-----NPLITNSL 539
DB 685 QQGAQLMTDYTVTSQFQNSLTHGANNRSMWYFLDTGYLNGDLINKKIVDKNRNSGTL 744
QY 540 VNRTDDNATAVPSPYSIRAHDSVQDLIADIKAIEINPNVVGY---SFTWEERIKKAFE 596
DB 745 VNRANSGETKVINYSFVRADHYDAQD---PIRKAMIDHGIKNMQDFTTFDQLAQOME 801
QY 597 IYNKOLLATE--KKYTHYNTALSAYLLTNKSSVPRVYVYGMFTDDGQYMAHKTINYBAI 654
DB 802 FYKQDENPSPGFKYNDYNLPSAYAMLTNKTQVPRVYVYGMFTDDGQYMEKGTIYNPVI 861
QY 655 ETLLKARIKYVSGGQAM-----NQOVNSEIITSVRYGKALKATDTGDRTT----- 702
DB 862 SALLKARIKYVSGGQMTATSDSGKLDGKGTDLTTSVRFGRGIM---TSQTTTQDQNSQ 917
QY 703 --RTSGVAVIEGNPRLKASDRVVVNMGAHKNQAVRPLLLTTDNGIKAYHSDQEAAG 760
DB 918 DYKQOGIIVIGNPDLKNNDKITLHMGKAKHKNQLYRALVLSNDSGLDVIDSDDKAPT 977
QY 761 LVRYTNRGELIFTAA-----DIKGYANPQVSGYGVVVPVGAALIKMFAIR 807
DB 978 L--RTNDNGDLIFHKNTNFVKQDGTIINYEMKGLSNALISGLVGVVVPVGAS----- 1027
QY 808 LARPHQOQVASY-----HQAALDSRVWFEGFSNFQAPATKKEEYTNVVIK 854
DB 1028 ---DSQDARTVATESSSSSNDGVSFHSNAALDSNVIEGFSNFQAMPPTSPEQSTNVVIATK 1084
QY 855 VDKFAEWGVTDFEMAPQVSVSDTG-----SFLDSVIONGYAFTDRYDLIGISK-----PN 903
DB 1085 ANLFKELGITSFELAPQVRSRSGDTNYGMSFLDSFLANGYAFTRDYDLGFKNADGNPPT 1144
QY 904 KYGTADDLVKAIKALHSKGIKVMADWPDQMYAFPEKEVVTATRVKDTGTPVAGSQIKNT 963
DB 1145 KYGTDDDLRNAIEALHKNQMAIADWPDQIYALPGKEVVTATRVDERGNQLKDTDFVNL 1204
QY 964 LYVVDGKSSGKQQAQYGAFLLELQAKYPELFAKQISTGVMPDPSVKIKQWSAKYFNG 1023

DB 1205 LYVANTKSSGDYQAKYGGFEFLDKLREYPSLFQKQNVSTQOPIDASTKIKQWSAKYFNG 1264
QY 1024 TNLIRGAGYVLKDOATNTYFNISDNKEINFLPKTLNQDSQVGSFYDGKGVYVYSTGY 1083
DB 1265 TNLIRGAGYVLKDOATNTYFNISDNKEINFLPKTLNQDSQVGSFYDGKGVYVYSTGY 1323
QY 1084 QAKTFTISEGD-KWYFDNNGYMWYGAQSGINGV-----NYFLSNGLOLRDAI 1130
DB 1324 QAKDTFIEDGNGNYFYDKOGYMYRSGQENPIRTVETSVNTRNGNYFMPNGVELRKGF 1383
QY 1131 LKNEDGTVAYVYVGNDRRYENGYQFMGSMWRHEN-NGEMSVGLTVIDGQ-VQYFDEMVGQ 1188
DB 1384 GTDSNGNYFYDDQGMVRDKYINDDANNFVHLNVDGMSRGLFKFSDTLOYPASNGVQ 1443
QY 1189 AKGFVTTADKIRYFKQSGNMYRNRPIENEEGK---W-----LYLGBDGAATVGSOTIN 1241
DB 1444 IKDSYAKDSKGNKYFDSATGN-----NDTGKATWDGNGYVITIDSDA-NNTIGVN 1494
QY 1242 GQHLVFRANGVQVKGFEVTDHHRISYYDNGSGD-----QIRNRFV-- 1282
DB 1495 TDYTAYITSSUREDLGFANAPYGVVT-KDQNGNDLKWOYINHTKQYEGQOVQVTRQYDTS 1553
QY 1283 -----RNAQGOFEYDNNNGYAVTCARTINGQLLYF---RANGVQVKGFEVTD 1326
DB 1554 KGVSNLITFAGGDLQGORLWDSRALTWTPTFKTMN-QISFISYANRNDGLFLNAPYQVK 1612
QY 1327 RYGRISYYDNGSGDQIRNRFVRNAQGO-WFYFDNNG---YAVTGARTINGQH---LYFR 1378
DB 1613 GYQLAGHSNQYKGOQVVIAGVANVSGKDWLSLSFNGTQWIDSQLNTNFTDMNQKVFV 1672
QY 1379 ANGVQVKGFEVTDHHRISYYDNGSGDQIRNRFVRNAQGOVWFYFDNNG---YAVTGART 1434
DB 1673 NTTSLNDGLFLNAPYROPYKLAGLAKNYYNQTVTVSQQ---YFDDQGTVMVSOVVLGGQT 1729
QY 1435 I-----NGOHLVFRANG 1446
DB 1730 VWVDNHALLAQMVSDTDOQLYVNSNG 1755

RESULT 11
US-09-604-957-3
; Sequence 3, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1278
; TYPE: PRT
; ORGANISM: Lactobacillus reuteri
US-09-604-957-3

Query Match 29.4%; Score 2276; DB 2; Length 1278;
Best Local Similarity 42.4%; Pred. No. 4.2e-144;
Matches 518; Conservative 183; Mismatches 382; Indels 140; Gaps 37;

QY 35 LVKADSNESKQIANDSNTSVVTANEESNVITEATSKOEAAASSQTN--HTVTTSSSTSV 92
DB 134 IVLVNGKEVRQLVND-----TKGAAAGFNRNDYVKNPATENSMS 174
QY 93 VNPKEVSNPVY-GETASNGEKLQNTTT-----VDKTSBAAAANNISKOTTEADTDVID 146

Db 175 SGFGIITLPTVTKNENVQLVHRFSNDVKTGEGNVYDFWSEL-----MPVKD 221
Qy 147 DSNAAQLILEKLPNVKEIDGKYIYD--NNGKVRNFTLIADGKILHFDGTGAYTDTSD 205
Db 222 SFQKNGPL--KQFGLQTINGQOYIIDPTTGPQRKNFLQSGNNWIYFD-----SDTGVG 274
Qy 206 TVNKDIV-----TTRSN-LYKKNQVYDRSAQSFHVHDHYLTAEWYRPKYILKDGKTWTQ 260
Db 275 TNALEQFAKGTVSSNEQYRNGNAAASYDDKSIENVNGYLTADTWYRPQILKDGTTWTD 334
Qy 261 STEKDFRPLMTWPDQETQRYVYNNAAQLGINKT-----YDDTSNQLNATAATIOAK 316
Db 335 SKETDMRPLMWPNLTQAYLYNMKQHGMLLPALPFFNADADPAELNHYSEIVQON 394
Qy 317 IEAKITTLKNTDWLRTQISAFVKTQSAMNSDSEK-PFDD-HLQNGAVLYDNEGKLTTPYAN 374
Db 395 IEKRISSETGNTDWLRTLMHDFVTNNPMKNKDSENVNFSGIQFGGFLKYEN-SDLTTPYAN 453
Qy 375 SNYRILNRTPTNQTGKDPRTADNTIGGYEFLANDVNSNPVQAEOLNWLHFLMNFQ 434
Db 454 SDYRLGRLMPIN-----IKDQTYRQGEFLANDIDNSNPVQAEOLNWLHYLLNFG 504
Qy 435 NIYANDPDANFDSIRVDADVNDADLLOIAGDYLAAGKIHKNDAANDHLSILEAWSN 494
Db 505 TITANNQANFDSVRVADPNIDADLNNIAQDYFNAAAYGM-DSDAVSNKHINILEDWHA 563
Qy 495 DTPYLHDDGNNMINMDNKLRLSLFLSAKLPLNORSNMPLITNSLVNRTDDNAETAAPS 554
Db 564 DPEYFNKIGNPOLTWDDTIKSNLHGLSDATN-RWGLDAIHQSADRENSTENVVPIN 622
Qy 555 YSFIKRAHSEVDPLADIADIIKAEINPNVVG---YSFTMEEIKKAFIYNKDLATEKKYTH 611
Db 623 YSFIKRAHSDNSQDQONAIR-----DVTGKYHTTFEDEQKIDAYIQDQNSTVKKYNL 677
Qy 612 YNTALSYALLLNKSSVRVYGDFTDDGQYMAHKTINYEAETILLKARIKYVSGQAM 671
Db 678 YNIPASYALLLNKDTIPRVYGDLYTDGQYMEHQTRYDYLTLNLLSRVKYVAGQOSM 737
Qy 672 RNQOV-GNSEIITSVRYKGALKAITDGTDRTRTSGVAVIEGNPNSLRKASDRVVNNG 730
Db 738 QTMSSVGGNNILTSVRYKGAMTATDTGTDETRTQIGVVSNTNFKLGVNDKVLHWG 797
Qy 731 AAHKNQAYRPLLTDTNGIKAYHSDOEAGLVRYTNDRGELIFTA-----AD--I 778
Db 798 AAHKNQAYRAAVLTITDGTGINTSYDQAP--VAMTDENGDLVLSHNLVNGKEADTAV 855
Qy 779 KYANPOVSYGLVGVVPGVGAALIKMFALEA---RPHQOMASVHONAAALDSRVMEGFSN 835
Db 856 QGYANPDVSGYLAVVVPVPGAS--DNQDARTAPSTKNSGNSAYRTNAAAPDSNVIFEAFSN 913
Qy 836 FOAPATKKEEYTNVIAKNVDKFAEWGVTDFEMAPQYVSSDTGSLFDSVIONGYAFTDRY 895
Db 914 FVYPTKESERANVRIAQNADEFASLGTFSFEMAPQYNSKDRTEFLDSTIDNGYAFTDRY 973
Qy 896 DLGISKPNKYGTADLVKAIKALHSGIKVMADWPQMYAPPEKEVVTATRVKDYKTPV 955
Db 974 DLGSEPNKYGTDEDLRNAIQALHKGALQVMADWPQDQYLYNLPGEVATVTRVDDRGVW 1033
Qy 956 AGSQIKNTLYVVDGKSSQDQAKYGGAFLEBLQAKYFELPARKOISTGVPMDPVSKIKQ 1015
Db 1034 KDAIINNNLYVN--TIGGGEYQKYGGAFLDLKQLYPEIFTKQVSTGVAIDPSOKITE 1092
Qy 1016 WSAKYFNGTNIILGRGAGVYLDQATNTYFNLSDNKEINFLPKTLNLDQO--VGSYDGK 1073
Db 1093 WSAKYFNGTNIILHRSGLGVFLKADG--GOYINLGTGTTK-QFLPLQLTGKQKQEGVEKNGD 1150
Qy 1074 G-YVYTSSTGYQAKNTFISEG-DKWWYFDNNGYMWYTGASIN-----GVNYYFLSGLQ 1125
Db 1151 GNYFYDLAGNMVKNVTFIEDSVGNWYFFDQDGKVENKHFDVDSYGEKGYFFLKNQVS 1210
Qy 1126 LRDAILKNEDGTAYIYGNDRRYENGYQYFMSGVWRHFNNGEMSVGLTVIDGQVOYFDEM 1185
Db 1211 FRGGLVOTDNGTY-YFDYNGKVRN-----QTINAGAM-----IYTLDEN 1249

Qy 1186 G-----YQAKGKFVTTAD-GKI 1201
Db 1250 GKLIKASYNDSAEYFTSTIDVGRM 1272

RESULT 12

US-09-995-749A-2
; Sequence 2, Application US/09995749A
; Patent No. 6867026
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: BOA3388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1781
; TYPE: PRT
; ORGANISM: Lactobacillus reuteri
US-09-995-749A-2

Query Match 29.4%; Score 2276; DB 2; Length 1781;
Best Local Similarity 42.4%; Pred. No. 6.9e-144;
Matches 518; Conservative 183; Mismatches 382; Indels 140; Gaps 37;

Qy 35 LVKADSNESKQISNDSNTSVVTANESNVITEATSKEAASSQTN--HTVTTSSSTSV 92
Db 637 IVLVNGKEVKGRLVND-----TKGAAGFRNDRVYKVNPAIENSSM 677
Qy 93 VNPKEVSNPYTV-GETASNGEKLQNTTT-----VDKTSEAAANNISKQTTEADTVD 146
Db 678 SGFGIITLPTVTKNENVQLVHRFSNDVKTGEGNVYDFWSEL-----MPVKD 724
Qy 147 DSNAAQLILEKLPNVKEIDGKYIYD--NNGKVRNFTLIADGKILHFDGTGAYTDTSD 205
Db 725 SFQKNGPL--KQFGLQTINGQOYIIDPTTGPQRKNFLQSGNNWIYFD-----SDTGVG 777
Qy 206 TVNKDIV-----TTRSN-LYKKNQVYDRSAQSFHVHDHYLTAEWYRPKYILKDGKTWTQ 260
Db 778 TNALEQFAKGTVSSNEQYRNGNAAASYDDKSIENVNGYLTADTWYRPQILKDGTTWTD 837
Qy 261 STEKDFRPLMTWPDQETQRYVYNNAAQLGINKT-----YDDTSNQLNATAATIOAK 316
Db 838 SKETDMRPLMWPNLTQAYLYNMKQHGMLLPALPFFNADADPAELNHYSEIVQON 897
Qy 317 IEAKITTLKNTDWLRTQISAFVKTQSAMNSDSEK-PFDD-HLQNGAVLYDNEGKLTTPYAN 374
Db 898 IEKRISSETGNTDWLRTLMHDFVTNNPMKNKDSENVNFSGIQFGGFLKYEN-SDLTTPYAN 956
Qy 375 SNYRILNRTPTNQTGKDPRTADNTIGGYEFLANDVNSNPVQAEOLNWLHFLMNFQ 434
Db 957 SDYRLGRLMPIN-----IKDQTYRQGEFLANDIDNSNPVQAEOLNWLHYLLNFG 1007
Qy 435 NIYANDPDANFDSIRVDADVNDADLLOIAGDYLAAGKIHKNDAANDHLSILEAWSN 494
Db 1008 TITANNQANFDSVRVADPNIDADLNNIAQDYFNAAAYGM-DSDAVSNKHINILEDWHA 1066
Qy 495 DTPYLHDDGNNMINMDNKLRLSLFLSAKLPLNORSNMPLITNSLVNRTDDNAETAAPS 554
Db 1067 DPEYFNKIGNPOLTWDDTIKSNLHGLSDATN-RWGLDAIHQSADRENSTENVVPIN 1125
Qy 555 YSFIKRAHSEVDPLADIADIIKAEINPNVVG---YSFTMEEIKKAFIYNKDLATEKKYTH 611

Db 1 LLANDIDNSNPVQAEQLNLHLYMNYGSI VANDPEANFDGVRVDAVDVNNADLLQIAD 60
Qy 467 YLKAAGIKHNDKAANDHLSILEAWSNDPTPYLLHDDGDNINNDKRLSLFLSLAKPLN 526
Db 61 YLKAHGVGDSEKNAINHLSILEAWSNDPOYNKDTKGALPDKRLSLFLALTRPLE 120
Qy 527 Q-----RSGMNPITNSLVNRTDDNAETAAPSYSFIRAHDSVQDLIADIKAETIN 578
Db 121 KDAASNKEIRSGLEPVITNSLNNRSAGKSERMANYIFIRAHDSVQTVIAKIIKAQIN 180
Qy 579 PNVGVGSFTMEELKKAFEIYNKDLALATEKYYTHYNTALSYALLTNKSSVPRVYGDMMFT 638
Db 181 PKTDGLTFLDELKQAFKTYNEDMRQAKKYYTQSNPTAYALMLSNKDSITRLYGDMS 240
Qy 639 DDQYMAHKTINVEAETLLKARIKYVSGQAMRNQOVGNSE-----IITSVRYCK 689
Db 241 DGOYQWATSPYDAIDTLLKARIKYAAGQDMKITVEGDKSHMDWDYTGVLTSVRYGT 300
Qy 690 GALKAFDTGDRTRTSGVAVIEGNPNSRLKASDRVVNMGAHKNQAYRPLLLTTDNGI 749
Db 301 GANEATDQSEATKTQGMVITSNPNSLKNQNDKVI VNMGAHKNQAYRPLLLTTKQGL 360
Qy 750 KAYHSDQEAAGLVRTNDRGELIFTAADIKG-YANPQVSGYLGWVPVGAALIKMFALRL 808
Db 361 TSYTSDAAKSLYRKTNDRGELVFDASDIQGLYNLPQVSG-LAVWVPVGAS--DNQDVRV 417
Qy 809 ARPHQOMAS---VHONAAALDSRMPEGSNFOAFATKEEYTNVIAKNVDFKFAEWGVT 865
Db 418 AASNKANATQVYESSALSDSLIYEGFSNFQDFTKSDYTNKIAQNVQVFKSGWVTS 477
Qy 866 FEMAPQYVSSDGSFLDSVIQNGYAFTDRYDLGISKPNKYGTADDLVKAIKALHSKGIKV 925
Db 478 FEMAPQYVSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSQDMINAVKALHKSQIQV 537
Qy 926 MADWVPDQ 933
Db 538 IADWVPDQ 545

RESULT 15
US-09-604-957-5
; Sequence 5, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 523
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-09-604-957-5

Query Match 20.0%; Score 1552; DB 2; Length 523;
Best Local Similarity 58.6%; Pred. No. 5.1e-96;
Matches 311; Conservative 76; Mismatches 132; Indels 12; Gaps 6;
Qy 407 LLANDVNSNPVQAEQLNLHLYMNYGSI VANDPEANFDGVRVDAVDVNNADLLQIAD 466
Db 1 LLANDVNSNPVQAEQLNLHLYMNYGSI VANDPEANFDGVRVDAVDVNNADLLQIAD 60
Qy 467 YLKAAGIKHNDKAANDHLSILEAWSNDPTPYLLHDDGDNINNDKRLSLFLSLAKPLN 526
Db 61 YLKAHGVGDSEKNAINHLSILEAWSNDPOYNKDTKGALPDKRLSLFLALTRPLE 120

Qy 527 QBSGMNPITNSLVNRTDDNAETAAPSYSFIRAHDSVQDLIADIKAETINPVV-GYS 585
Db 121 IEGTWQRFVYDMVDRSNDSTENEALPNYSFVRAHDSVQTVIAQIV-SOLYDPDENSLSA 179
Qy 586 FTMEIKKAFEIYNKDLALATEKYYTHYNTALSYALLTNKSSVPRVYVYGDMMFTDDGQYMA 645
Db 180 PTTEQLAAAFKVYNEDEKLADKKYTOYNMASAYAMLLTNKDTVPRVYVYGDLYTDDGQYMA 239
Qy 646 HKTINYEAEITLLKARIKYVSGQAMRNQOVGNSEIITSVRYGKGALKATDGTDRTRTS 705
Db 240 TKSPYDAIDTLLKARVQVYAGQSN---SVDSNDVLTSVRYGKDMATSDTGTSETRTE 296
Qy 706 GVAVIEGNPNSURLKASDRVVNMGAHKNQAYRPLLLTTDNGIKAYHSDQEAAGLVRYT 765
Db 297 GIGVIVSNNAEQLQEDGHTVTLHMGAAHKNQAYRALLSTADGLAYYDTDENAP--VAYT 354
Qy 766 NDRGELIFTAADIKGYPANPQVSGYLGWVPVGAALIKMFALRLARPHQOMAS---VHONA 822
Db 355 DANGDLIFTNESIYGVQNPQVSGYLAVWVPVGAQDQ--DARTASPTTTNTSDKVPHSNA 412
Qy 823 ALDSRMPEGSNFOAFATKEEYTNVIAKNVDFKFAEWGVTDFEMAPQYVSSDGSFLD 882
Db 413 ALDSQVIYEGFSNFOAFATDSSEYTNVIAQNAQDQFKQMGVTSFQLAPQYRSSTDTSF 472
Qy 883 SVIQNGYAFTDRYDLGISKPNKYGTADDLVKAIKALHSKGIKVMADWVPDQ 933
Db 473 SIIQNGYAFTDRYDLGYSPTKYGTADQLRDAIKALHASGIQAIADWVPDQ 523

Search completed: February 11, 2006, 19:45:34
Job time: 38.2941 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 20:39:23 ; Search time 132.022 Seconds
(without alignments)
4668.162 Million cell updates/sec

Title: US-10-797-821-34
Perfect score: 7741
Sequence: 1 MDKKVRYKLRVKRWTVS.....DRYGRISYYDANSRGRVIRN 1475

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA_Main:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7741	100.0	1475	3	US-09-740-274-2
2	7741	100.0	1475	4	US-10-383-930-34
3	7741	100.0	1475	5	US-10-797-821-34
4	5285	68.3	1375	3	US-09-740-274-4
5	5285	68.3	1375	4	US-10-383-930-35
6	5285	68.3	1375	5	US-10-797-821-35
7	4580	59.2	1590	4	US-10-383-930-37
8	4580	59.2	1590	5	US-10-797-821-37
9	3817.5	49.3	1430	3	US-09-740-274-6
10	3817.5	49.3	1430	4	US-10-383-930-36
11	3817.5	49.3	1430	5	US-10-797-821-36
12	3434.5	44.4	1554	4	US-10-383-930-38
13	3434.5	44.4	1554	5	US-10-797-821-38
14	3196.5	41.3	1518	4	US-10-383-930-40
15	3196.5	41.3	1518	5	US-10-797-821-40
16	3161	40.8	1497	5	US-10-484-218-18
17	2945.5	38.1	1365	4	US-10-383-930-39
18	2945.5	38.1	1365	5	US-10-797-821-39
19	2877.5	37.2	1595	5	US-10-484-218-20
20	2537	32.8	2057	4	US-10-417-280A-2
21	2460	31.8	1006	5	US-10-484-218-22
22	2340	30.2	1777	5	US-10-484-218-12
23	2321	30.0	1771	5	US-10-484-218-14
24	2276	29.4	1781	3	US-09-995-749A-2
25	1657	21.4	545	3	US-09-995-749A-10
26	1547.5	20.0	522	3	US-09-995-749A-11
27	1379	17.8	535	3	US-09-995-749A-13

28	1377.5	17.8	787	5	US-10-484-218-16	Sequence 16, Appl
29	1289.5	16.7	584	3	US-09-995-749A-12	Sequence 12, Appl
30	1249.5	16.1	525	5	US-10-484-218-23	Sequence 23, Appl
31	683.5	8.8	224	5	US-10-484-218-4	Sequence 4, Appl
32	673	8.7	223	5	US-10-484-218-6	Sequence 6, Appl
33	647	8.4	223	5	US-10-484-218-10	Sequence 10, Appl
34	577	7.5	221	5	US-10-484-218-2	Sequence 2, Appl
35	552	7.1	221	5	US-10-484-218-8	Sequence 8, Appl
36	456	5.9	2710	4	US-10-011-366-6	Sequence 6, Appl
37	456	5.9	2710	4	US-10-354-774-6	Sequence 6, Appl
38	456	5.9	2710	4	US-10-271-012-6	Sequence 6, Appl
39	456	5.9	2710	4	US-10-729-122-6	Sequence 6, Appl
40	456	5.9	2710	4	US-10-729-039-6	Sequence 6, Appl
41	456	5.9	2710	5	US-10-729-527-6	Sequence 6, Appl
42	456	5.9	2710	5	US-10-727-898-6	Sequence 6, Appl
43	456	5.9	2710	6	US-10-728-696-6	Sequence 6, Appl
44	456	5.9	2710	6	US-11-001-241-6	Sequence 6, Appl
45	402	5.2	811	4	US-10-011-366-7	Sequence 7, Appl

ALIGNMENTS

RESULT 1
US-09-740-274-2
; Sequence 2, Application US/09740274
; Patent No. US20020031826A1
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans

Query Match	100.0%	Score	7741;	DB	3;	Length	1475;
Best Local Similarity	100.0%	Pred. No.	0;				
Matches	1475;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	MDKKVRYKLRVKRWTVSVASAVMTLTTLTSSGGLVKADSNESKSQISNDSNTSVVTANE	60				
Db	1	MDKKVRYKLRVKRWTVSVASAVMTLTTLTSSGGLVKADSNESKSQISNDSNTSVVTANE	60				
Qy	61	ESNVITEATSKQEAASQTNNHTVTTSSSTSVVNPKEVVSNPYTVGTETASNGEKLQNTT	120				
Db	61	ESNVITEATSKQEAASQTNNHTVTTSSSTSVVNPKEVVSNPYTVGTETASNGEKLQNTT	120				
Qy	121	TVDKTSEAAANNISKQTTQTEADTDVDSNAAANLQLEKLPNVKEIDGKYVYDNNKGVRT	180				
Db	121	TVDKTSEAAANNISKQTTQTEADTDVDSNAAANLQLEKLPNVKEIDGKYVYDNNKGVRT	180				
Qy	181	NFTLIADGKILHFDETGAITDTSITVKNKOITVTRSNLYKKYQNVYDRSAQSFEHVDHYL	240				
Db	181	NFTLIADGKILHFDETGAITDTSITVKNKOITVTRSNLYKKYQNVYDRSAQSFEHVDHYL	240				

Db 181 NPTLIADGKILHFDEGTAYDTSIDTVANKD I V T T R S N L Y K K Y N Q V Y D R S A Q S F E H V D H Y L 240

Qy 241 T A E S W Y R P K Y I L K D G K T W T Q S T E K D F R P L L M T W P D Q E T Q R O Y Y N Y M N A Q L G I N K T Y D D T 300

Db 241 T A E S W Y R P K Y I L K D G K T W T Q S T E K D F R P L L M T W P D Q E T Q R O Y Y N Y M N A Q L G I N K T Y D D T 300

Qy 301 S N Q L Q N I A A A T I O A K I E A K I T T L K N T D W L R O T I S A F V K T O S A M N S D S E K P D D H L O N G A 360

Db 301 S N Q L Q N I A A A T I O A K I E A K I T T L K N T D W L R O T I S A F V K T O S A M N S D S E K P D D H L O N G A 360

Qy 361 V L Y D N E G K L T P Y A N S N Y R I L N R T P N T Q T G K D P R Y T A D N T I G G Y E F L L A N D V D N S N P V V Q 420

Db 361 V L Y D N E G K L T P Y A N S N Y R I L N R T P N T Q T G K D P R Y T A D N T I G G Y E F L L A N D V D N S N P V V Q 420

Qy 421 A E Q L N W L H F L M N F G N I Y A N D P A N F D S I R V D A V D N V D A D L L Q I A G D Y L K A A G I H K N D K A 480

Db 421 A E Q L N W L H F L M N F G N I Y A N D P A N F D S I R V D A V D N V D A D L L Q I A G D Y L K A A G I H K N D K A 480

Qy 481 A N D H L S I L E A W S D N D T P V L H D D G N M I N M D N K L R L S L F S L A K P L N Q R S G M N P L I T N S I V 540

Db 481 A N D H L S I L E A W S D N D T P V L H D D G N M I N M D N K L R L S L F S L A K P L N Q R S G M N P L I T N S I V 540

Qy 541 N R T D N A E T A A V P S Y S F I R A H D S E V Q D L I A D I I K A E I N P N V V G Y S F T M E E I K K A F E I Y N K 600

Db 541 N R T D N A E T A A V P S Y S F I R A H D S E V Q D L I A D I I K A E I N P N V V G Y S F T M E E I K K A F E I Y N K 600

Qy 601 D L L A T E K K Y T H Y N T A L S V A L L T N K S S P R V Y Y G M F T D D G Y M A H K T I N Y E A I E T L L K A 660

Db 601 D L L A T E K K Y T H Y N T A L S V A L L T N K S S P R V Y Y G M F T D D G Y M A H K T I N Y E A I E T L L K A 660

Qy 661 R I K Y S G G O A M N Q O V G N S E I T S V R Y G K G A L K A T D T G D R T R T S G V A V I E G N N P S L R L K 720

Db 661 R I K Y S G G O A M N Q O V G N S E I T S V R Y G K G A L K A T D T G D R T R T S G V A V I E G N N P S L R L K 720

Qy 721 A S D R V V N M G A A H K N Q A Y R P L L L T T D N G I K A Y H S Q E A A G L V R Y T N D R G E L I F T A A D I K G 780

Db 721 A S D R V V N M G A A H K N Q A Y R P L L L T T D N G I K A Y H S Q E A A G L V R Y T N D R G E L I F T A A D I K G 780

Qy 781 Y A N P O V S G Y L G V W P V G A A L I K M F A L R L A R P H Q O M A S V H Q N A A L D S R V M F E G S F N Q A P A 840

Db 781 Y A N P O V S G Y L G V W P V G A A L I K M F A L R L A R P H Q O M A S V H Q N A A L D S R V M F E G S F N Q A P A 840

Qy 841 T K K E Y T N V I T A K N V D K F A E G V T D F E M A P Q V Y S T D G S F L D S V I O N G Y A F T D R Y D L G I S 900

Db 841 T K K E Y T N V I T A K N V D K F A E G V T D F E M A P Q V Y S T D G S F L D S V I O N G Y A F T D R Y D L G I S 900

Qy 901 K P N K Y G T A D D L V K A I K A L H S K G I K V M A D W P D M Y A F P E K E V V T A T R V D K Y G T P V A G S Q I 960

Db 901 K P N K Y G T A D D L V K A I K A L H S K G I K V M A D W P D M Y A F P E K E V V T A T R V D K Y G T P V A G S Q I 960

Qy 961 K N T L Y V D G K S G K D Q Q A K Y G A F L E E L O A K Y P E L F A R K Q I S T G V P M D P S V K I K O W S A K Y 1020

Db 961 K N T L Y V D G K S G K D Q Q A K Y G A F L E E L O A K Y P E L F A R K Q I S T G V P M D P S V K I K O W S A K Y 1020

Qy 1021 F N G T N I L G R G A G Y L K D Q A T N Y F N I S D N K E I N F L P K T L L N Q D S O V G S Y D G K G V Y Y Y S T 1080

Db 1021 F N G T N I L G R G A G Y L K D Q A T N Y F N I S D N K E I N F L P K T L L N Q D S O V G S Y D G K G V Y Y Y S T 1080

Qy 1081 S G Y Q A K N T F I S E G D K W Y Y F D N N G Y M V T G A Q S I N G V N Y Y F L S N G L Q R D A I L K N E D G T Y A Y 1140

Db 1081 S G Y Q A K N T F I S E G D K W Y Y F D N N G Y M V T G A Q S I N G V N Y Y F L S N G L Q R D A I L K N E D G T Y A Y 1140

Qy 1141 Y G N D G R R Y E N G Y Q P M S G W R H F N N G E M S V G L T V I D G V Q V P D E M G Y O A K G K F V T T A D G K 1200

Db 1141 Y G N D G R R Y E N G Y Q P M S G W R H F N N G E M S V G L T V I D G V Q V P D E M G Y O A K G K F V T T A D G K 1200

Qy 1201 I R Y P D K O S G N M Y R N F I E N E E G K W I Y L G E D G A A V T G S O T I N G O H L Y F R A N G V Q V K G E F V T 1260

Db 1201 I R Y P D K O S G N M Y R N F I E N E E G K W I Y L G E D G A A V T G S O T I N G O H L Y F R A N G V Q V K G E F V T 1260

Qy 1261 D H H G R I S Y Y D G N S G Q I R N R F V R N A Q O G W F Y D N N G Y A V T G A R T I N G Q L L Y F R A N G V Q V K 1320

Db 1261 D H H G R I S Y Y D G N S G Q I R N R F V R N A Q O G W F Y D N N G Y A V T G A R T I N G Q L L Y F R A N G V Q V K 1320

RESULT 2

US-10-383-930-34
; Sequence 34, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 34
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-34

Query Match 100.0%; Score 7741; DB 4; Length 1475;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1475; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 M D K K V R Y K L R V K K R W T V S V A S A V M T L T L S G G L V K A D S N E S K S Q I S N D S N T S V V T A N E 60

Db 1 M D K K V R Y K L R V K K R W T V S V A S A V M T L T L S G G L V K A D S N E S K S Q I S N D S N T S V V T A N E 60

Qy 61 E S N V I T E A T S K Q E A A S S Q T N H T V T T S S S T S V N P K E Y V S N P Y T V G E T A S N G E K L O N Q T T 120

Db 61 E S N V I T E A T S K Q E A A S S Q T N H T V T T S S S T S V N P K E Y V S N P Y T V G E T A S N G E K L O N Q T T 120

Qy 121 T V D K T S E A A A N N I S K Q T T E A D T D V I D D S N A A N L Q I L E K L P N V K E I D G K Y Y Y D N N G K V R T 180

Db 121 T V D K T S E A A A N N I S K Q T T E A D T D V I D D S N A A N L Q I L E K L P N V K E I D G K Y Y Y D N N G K V R T 180

Qy 181 N F T L I A D G K I L H F D E T G A Y T D T S I D T V N K D I V T T R S N L Y K K Y N Q V Y D R S A Q S F E H V D H Y L 240

Db 181 N F T L I A D G K I L H F D E T G A Y T D T S I D T V N K D I V T T R S N L Y K K Y N Q V Y D R S A Q S F E H V D H Y L 240

Qy 241 T A E S W Y R P K Y I L K D G K T W T Q S T E K D F R P L L M T W P D Q E T Q R O Y Y N Y M N A Q L G I N K T Y D D T 300

Db 241 T A E S W Y R P K Y I L K D G K T W T Q S T E K D F R P L L M T W P D Q E T Q R O Y Y N Y M N A Q L G I N K T Y D D T 300

Qy 301 S N Q L Q N I A A A T I O A K I E A K I T T L K N T D W L R O T I S A F V K T O S A M N S D S E K P D D H L O N G A 360

Db 301 S N Q L Q N I A A A T I O A K I E A K I T T L K N T D W L R O T I S A F V K T O S A M N S D S E K P D D H L O N G A 360

Qy 361 V L Y D N E G K L T P Y A N S N Y R I L N R T P N T Q T G K D P R Y T A D N T I G G Y E F L L A N D V D N S N P V V Q 420

Db 361 V L Y D N E G K L T P Y A N S N Y R I L N R T P N T Q T G K D P R Y T A D N T I G G Y E F L L A N D V D N S N P V V Q 420

Qy 421 A E Q L N W L H F L M N F G N I Y A N D P A N F D S I R V D A V D N V D A D L L Q I A G D Y L K A A G I H K N D K A 480

Db 421 A E Q L N W L H F L M N F G N I Y A N D P A N F D S I R V D A V D N V D A D L L Q I A G D Y L K A A G I H K N D K A 480

```
QY 481 ANDHLSILEAWSNDTPYLHDDGDNMNMNDKRLSLFSLAKPLNQRSGMPLTNSLV 540
Db 481 ANDHLSILEAWSNDTPYLHDDGDNMNMNDKRLSLFSLAKPLNQRSGMPLTNSLV 540
QY 541 NRTDDNAETAAPVPSYFIRAHSEVQDLADIKAIEINPNVVGYSFTMEEIKKAEIYNK 600
Db 541 NRTDDNAETAAPVPSYFIRAHSEVQDLADIKAIEINPNVVGYSFTMEEIKKAEIYNK 600
QY 601 DLLATEKKYTHYNTALSAYALLTNKSSVPRVYVYGDMDFTDDGVMMAHKTINYEAIETLLKA 660
Db 601 DLLATEKKYTHYNTALSAYALLTNKSSVPRVYVYGDMDFTDDGVMMAHKTINYEAIETLLKA 660
QY 661 RIKYVSGQAMRNQVGNSEIITSVRYGKALKATDGTDRTRTSGVAVIEGNPRLRK 720
Db 661 RIKYVSGQAMRNQVGNSEIITSVRYGKALKATDGTDRTRTSGVAVIEGNPRLRK 720
QY 721 ASDRVVNVNNGAAHKNQAVPEPLLTDTNGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKG 780
Db 721 ASDRVVNVNNGAAHKNQAVPEPLLTDTNGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKG 780
QY 781 YANPOVSGVLGVVPGVGAALIKWFAALRLARPHQQWASVHONAAALSRVMFEGFSFQAPA 840
Db 781 YANPOVSGVLGVVPGVGAALIKWFAALRLARPHQQWASVHONAAALSRVMFEGFSFQAPA 840
QY 841 TKKEEYTNVVIKQVNDKFAEWGVTDFEMAPQVVSSTGDSFLDSVIQNGVAFTRDYLGIS 900
Db 841 TKKEEYTNVVIKQVNDKFAEWGVTDFEMAPQVVSSTGDSFLDSVIQNGVAFTRDYLGIS 900
QY 901 KENKYGTADDLKAIKALHSKIGIKWADVPQWAFPEKEVVTATRVDKYGTTPVAGSQI 960
Db 901 KENKYGTADDLKAIKALHSKIGIKWADVPQWAFPEKEVVTATRVDKYGTTPVAGSQI 960
QY 961 KNTLVVDGKSGKQQAQYGAFLLELOAKYPELPARKQIISTGVPMDBPSVKIKOWSAKY 1020
Db 961 KNTLVVDGKSGKQQAQYGAFLLELOAKYPELPARKQIISTGVPMDBPSVKIKOWSAKY 1020
QY 1021 FNGTNILGRGAGYVLKQATNTYFNISDNKEINFLPKTLNLNDSQVGFSDGKGVYVYST 1080
Db 1021 FNGTNILGRGAGYVLKQATNTYFNISDNKEINFLPKTLNLNDSQVGFSDGKGVYVYST 1080
QY 1081 SGYQAKNTFISEGDKWYYPDNNGYVMTGAQSIINGVNYPLSLNGLQRLDAILKNEDGTAY 1140
Db 1081 SGYQAKNTFISEGDKWYYPDNNGYVMTGAQSIINGVNYPLSLNGLQRLDAILKNEDGTAY 1140
QY 1141 YGNDGRRYENGYYQFMGQWVRHFNNGEMSVGLTVIDGQVQYFDEMGYQAKGKFTVTDGK 1200
Db 1141 YGNDGRRYENGYYQFMGQWVRHFNNGEMSVGLTVIDGQVQYFDEMGYQAKGKFTVTDGK 1200
QY 1201 IRYFDKQSGNMYRNRFIENEKGWLYLGEDGAAVTGSOTINGHLYFRANGVQVKGFTV 1260
Db 1201 IRYFDKQSGNMYRNRFIENEKGWLYLGEDGAAVTGSOTINGHLYFRANGVQVKGFTV 1260
QY 1261 DHGGRISYDNGSGDQIRNRFVRNAQGWYFDDNNGYAVTGARTINGQLLYFRANGVQVK 1320
Db 1261 DHGGRISYDNGSGDQIRNRFVRNAQGWYFDDNNGYAVTGARTINGQLLYFRANGVQVK 1320
QY 1321 GEFVTDYGRISYDNGSGDQIRNRFVRNAQGWYFDDNNGYAVTGARTINGHLYFRAN 1380
Db 1321 GEFVTDYGRISYDNGSGDQIRNRFVRNAQGWYFDDNNGYAVTGARTINGHLYFRAN 1380
QY 1381 GVQVKGFEVTDYGRISYDNGSGDQIRNRFVRNAQGWYFDDNNGYAVTGARTINGOHL 1440
Db 1381 GVQVKGFEVTDYGRISYDNGSGDQIRNRFVRNAQGWYFDDNNGYAVTGARTINGOHL 1440
QY 1441 YFRANGVQVKGFEVTDYGRISYDNGSGDQIRNRFVRNAQGWYFDDNNGYAVTGARTINGOHL 1475
Db 1441 YFRANGVQVKGFEVTDYGRISYDNGSGDQIRNRFVRNAQGWYFDDNNGYAVTGARTINGOHL 1475
```

RESULT 3

US-10-797-821-34

; Sequence 34, Application US/10797821

; Publication No. US20050031633A1

; GENERAL INFORMATION:

```
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 34
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
; US-10-797-821-34
```

Query Match 100.0%; Score 7741; DB 5; Length 1475;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1475; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MDKVRVYKLRVKKRWTVSVASAVMTLTTLTSGGLVKADSNESKQISNDSNTSVVTANE 60
Db 1 MDKVRVYKLRVKKRWTVSVASAVMTLTTLTSGGLVKADSNESKQISNDSNTSVVTANE 60
QY 61 ESNVITEATSKQEAASSQTNHTVTTSSTSVNPKVSVNPPYTVGETASNGEKLQNTT 120
Db 61 ESNVITEATSKQEAASSQTNHTVTTSSTSVNPKVSVNPPYTVGETASNGEKLQNTT 120
QY 121 TVDKTSEAAANNISKQTTTEADTVLDDSNAAHLQLEKLPNVKEIDGKYIYDNNKGVRT 180
Db 121 TVDKTSEAAANNISKQTTTEADTVLDDSNAAHLQLEKLPNVKEIDGKYIYDNNKGVRT 180
QY 181 NPTLIADKILHFDETGAYTDTSDTVNKLIVTTRSNLYKKYNOVYVDRSAQSFEHVDHYL 240
Db 181 NPTLIADKILHFDETGAYTDTSDTVNKLIVTTRSNLYKKYNOVYVDRSAQSFEHVDHYL 240
QY 241 TAESWYRPKYILKDGKWTQSTEKDFRPLMTWMPDQETQRYVNMNAQLGINKTYDDT 300
Db 241 TAESWYRPKYILKDGKWTQSTEKDFRPLMTWMPDQETQRYVNMNAQLGINKTYDDT 300
QY 301 SNQLNLNTAAATIQAIEAKITTLKNTDWLRQTTISAFVKTQSAMNSDSEKPPDDHLONGA 360
Db 301 SNQLNLNTAAATIQAIEAKITTLKNTDWLRQTTISAFVKTQSAMNSDSEKPPDDHLONGA 360
QY 361 VLYDNEGKLTTPYANSNYRILNRTPTNQTGKDPRTVADNTTIGGYEFLANDVDNPNPVQ 420
Db 361 VLYDNEGKLTTPYANSNYRILNRTPTNQTGKDPRTVADNTTIGGYEFLANDVDNPNPVQ 420
QY 421 AEQLNWLHFLMNFNFIYANDPDANFDSIRVDAVNVDADLLQIAGDYLKAAKGIHKNDKA 480
Db 421 AEQLNWLHFLMNFNFIYANDPDANFDSIRVDAVNVDADLLQIAGDYLKAAKGIHKNDKA 480
QY 481 ANDHLSILEAWSNDTPYLHDDGDNMNMNDKRLSLFSLAKPLNQRSGMPLTNSLV 540
Db 481 ANDHLSILEAWSNDTPYLHDDGDNMNMNDKRLSLFSLAKPLNQRSGMPLTNSLV 540
QY 541 NRTDDNAETAAPVPSYFIRAHSEVQDLADIKAIEINPNVVGYSFTMEEIKKAEIYNK 600
Db 541 NRTDDNAETAAPVPSYFIRAHSEVQDLADIKAIEINPNVVGYSFTMEEIKKAEIYNK 600
QY 601 DLLATEKKYTHYNTALSAYALLTNKSSVPRVYVYGDMDFTDDGVMMAHKTINYEAIETLLKA 660
Db 601 DLLATEKKYTHYNTALSAYALLTNKSSVPRVYVYGDMDFTDDGVMMAHKTINYEAIETLLKA 660
```



```
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRP
; ORGANISM: streptococcus mutans
US-09-740-274-6

Query Match      49.3%; Score 3817.5; DB 3; Length 1430;
Best Local Similarity 52.9%; Pred. No. 9e-212;
Matches 762; Conservative 231; Mismatches 394; Indels 53; Gaps 22;

QY 1 MDKVRYYKLRYKVKRWVTVSVASAVMTL--TTLGGGLVKADSNESKS-----QISNDSNTS 54
Db 1 METKRRYKMKVKRWVTVSVASAVMTL--TTLGGGLVKADSNESKS-----QISNDSNTS 60

QY 55 VVTANEESNVITEATS KQBAASSQTNHHTVTTSSSSTSVVNPKEVSNPYTVGTETASNGEK 114
Db 61 SESSQTDAPKTQAOEQTEQTAQSOAN--VADTSTSTIKETPSQNIITTOANSDDKTVTNYS 119

QY 115 LQNTTIV--DKTSEAAANNISKOT--TEADTVIDDSNAA-----NLQILEKLPVNYKEIDG 167
Db 120 BEAQTSEERTKQSEEAQTASSQALTKQAKELTKQRTAAQENKPNVDLAAIPNVKQIDG 179

QY 168 KYYYDNNKVRNTETLADGKILHFE--TGAYTDTSDTIDVTKNDIVTTRSNLYKKYNQY 226
Db 180 KYYYIGSDGQPKKFNALTVNNKVFYDKNTGALTDTTS--QYQPKQGLTKLNDYTPHNOIV 238

QY 227 DRSAQSFHVDHYLTAESWYRPKYILKDGKWTQTQTEKDFRLLMTWPDQETQRYVNY 286
Db 239 NFENTSLETIDNYVADS WYRPKPDILKNGKWTWASESDLRPLMSWPFDKQTQIAYLNY 298

QY 287 MNAQ--LGINKTVDDTSNQLNIAAATIQAIEAKITTLKNTDMLRQTISAFVKTQSAWN 345
Db 299 MNQOGLGTGENYADSSQESLNLAQTQVKIETKISQTOQTQWLRIINSFVKTPQPNW 358

QY 346 SDSEKPFDD----DHLONGAVLYDNEGKLTTPYANSYRILNRPPTNQTGKDKPRIYADNTI 401
Db 359 SQTESDTSAGEKHQGGGALLYNSDK--TAYANSYRILNRPPTTSQTGK--PKYPEDNSS 415

QY 402 GGYEFLLANDVDSNPVQAEQLNLHFLNFGNIYANDPDANFDSIRVDAVDNVDADLL 461
Db 416 GGYDFLLANDVDSNPVQAEQLNLHFLNFGNIYANDPDANFDSIRVDAVDNVDADLL 475

QY 462 QIAGDYLLKAAKGIHKNKAANDHLSILEAWSNDPTPYLHDDGDNMINNDKRLSLFLSL 521
Db 476 QIASDYLLKAAHYGVDSKSEKNAIHLHLSILEAWSNDPDQYNKDTKGAQLPIDNKLRLSLLYAL 535

QY 522 AKPLNQ-----RSGNPLITNSLVNRTDNDNAETAAPVSPSIFRAHDSVQDLADI 573
Db 536 TRPLEKDSAKNEIRISGLEPVTNSLNNRSABGKNSERMANYIFIRAHDSVQTVIAKII 595

QY 574 KAEINPNVVGVSFTMEELKKAPEIYNKDLATEKKYTHYNTALSVALLLTNKSSVPRVY 633
Db 596 KAIINPKTDGLTFTLDEULKAQAFKIYEDNRQAKKKTQSNIPTAYALMLSNKDSITRLYY 655

QY 634 GDMFTDDGQYMAHKTINYEAEITLLKARIKYVSGGAAMNQOVGNSE-----IITS 684
Db 656 GDMYSDDGQYMATKSPYDAIDTLKARIKAAAGGQDMKITYVEGDKSHMDWDYTVGLTS 715

QY 685 VRYGKALKATDGTGRTTTSVAVIEGNNPSLRLLKASDRVVVNNMGAHKNQAYPLLLT 744
Db 716 VRYGTGANEATDQGSSEATKTQGMVITSNPPLKLNQNDKVIIVNNMGAHKNQAYPLLLT 775
```

```
QY 745 TDNGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKGYANPQVSGYLVVVPVGAALIKMF 804
Db 776 TKDGLTSYTSDAAKSLYKTKNDKGELVFDASDIQGYLNPQVSGYLVVVPVGAS--DNQ 833

QY 805 ALRLARPHQOMAS---VHQNAALDSRVMEFEGSNFOAFATKKEEYTNVVIKNDVKPAEW 861
Db 834 DVRVAASNKANATGQVYESSALDSQLIYEGFSNFQDFVTKDSYTNKKIAQNVQLFKSW 893

QY 862 GYTDPEMAPQVYSSYTDGSLDSVIONGYAFTDRYDLGISGPNKYGTGADDLVKAIKALHSK 921
Db 894 GVTSPFEMAPQVYSSSDGSLDSIIQNGYAFEDRYDLAMSKNNKYSQQQDMINAVKALHSK 953

QY 922 GIKVMADWVPDQMYAFPEKVEVVTATRVDKYGTVPVAGSQIKNTLYVVDGKSSGKDOQAKY 981
Db 954 GIQVIADWVPDQIYNLPKKEVVTATRVNDYGEYRKDSEIKNTLYAANTKSNKGDYQAKY 1013

QY 982 GAFLELOAKYPELPFARKQISTGVPMWDPVSKIKQNSAKYFNGTNIILGRAGYVYLKQOATN 1041
Db 1014 GAFLELOAKYPSIFNRTQISNGKKIDPSEKITAWKAKYFNGTNIILGRAGYVYLKQOATN 1073

QY 1042 TYFNIISDNKEINFLPKTLNODSOVGFSDYDGKGVYVYSTGYQAKNTEISEG--DKWY 1100
Db 1074 KYFELKGNQ--TYLPKQMTNKEASTGFVNDGNGMTIFYSTGYQAKNSFYQDAKGNWYFD 1131

QY 1101 NNGYVMTGAQSIINGVNYIFLSNGLQLRDAILKNEDGTAYAYGNDGRRYENGYYQF--MSGV 1159
Db 1132 NNGHMYVGLQQLNGEVQVFLSNGVQLRESFLENADGSKNYFGLHGNRYNGYYSFNDK 1191

QY 1160 WRHFN--NEMSVMGLTVIDGQVQVDEMGMQYQAKGKFTTADGKIRYFDKOSGNMYNRFTIE 1218
Db 1192 WRYFDASGVMAVGLKTINGTQYFDQDGQYQVKGAWITGSDGKKRYFDGSGNMAVNRVAN 1251

QY 1219 NEEGKWLVLGDEGAATVGSOTINGOHLFRANGVOVKGEFVTDHGRISYDNGSGDQIR 1278
Db 1252 DKNGDWYLYNSDGIALVGVTINGKTYTFYFGQDGKQIKGLIITD--NGKLYFLANS 1310

QY 1279 NRVFVNAQCFYFDNNGYAVTGTARTINGQLLYFRANGVQVKGEFVTDYGRISYYDNGS 1338
Db 1311 NIFATDSQNNMYFGSDGVAVTGSQTIAKKLYFASDGKQVKGSPVTYN--GKVVHYHADS 1369

QY 1339 GDQIINRVFVNAQCFYFDNNGYAVTGTARTINGOHLFRANGVQVKGEFVTDHGRISY 1398
Db 1370 GELQVNRFEADKGNWYLYDSNGEALTSQRIINDQVFFTRREGKQVKGVDAYDERLLVY 1429

RESULT 10
US-10-383-930-36
; Sequence 36, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 36
; LENGTH: 1430
; TYPE: PRP
; ORGANISM: Streptococcus mutans
US-10-383-930-36

Query Match      49.3%; Score 3817.5; DB 4; Length 1430;
Best Local Similarity 52.9%; Pred. No. 9e-212;
Matches 762; Conservative 231; Mismatches 394; Indels 53; Gaps 22;
```



```
Db 180 KYIIGSDGQPKNFALTVNNKVLVYFDKNTGALTDTIS-QYQFKQGLTKLNDVTPINQIV 238
QY 227 DRSAOSFEHVDHYLTAEISWYRPKYILLKOGKWTQTSTEKDFRPLPLMTWPDQETQROYVNY 286
Db 239 NFENTSLETIDNVYVADSWYRPKIDILKNGKWTATASSESDLRPLPLMSWPDKQTIAYLNY 298
QY 287 KNAQ-LGINKTYDDTSNQLQAINAAATIAQAEAKITTLKNTDNLBQTTISAFVKTOSAWN 345
Db 299 MNQQLGTGENTYADSSQBSLAAQTQVQKIEKISQOTQOTQWLRIINSEVKTQPNWN 358
QY 346 SDSEKPFDD---DHLQNGAVLYDNEGLTPYANSNYRIILNRTPTNQTGKDKPRYADNTI 401
Db 359 SQTESDTSAGEKHDLGGGALLYSNSDK-TAYANSDYRLNRTPTSTQTK--PKYFEDNS 415
QY 402 GGYEFLLANDVNSNPVQAEQLNLHFLMNFNIYANDPDANFDSIRVDVNDVADALL 461
Db 416 GGYDFLLANDIDNSNPVQAEQLNLHFLMNYGSIIVANDPEANFDGVRVDAVDNVNADLL 475
QY 462 QIAGDYLLKAAKGHIKNDKAAANDHLSLILEAWSNDPTLYLHDDGDNMINMDNKLRLSLFSL 521
Db 476 QIASDYLLKAAHYGVDSKNAIINHLSLILEAWSNDPQYNKDTKGAQLPIDNKLRLSLLYAL 535
QY 522 AKPLAQ-----RSGMPLITNSLVNRTDDNAETAAPVSYSFIRAHSEVQDLIADI 573
Db 536 TRPLEKDSAKNEIRSGLEPVITNSLNNRSAGKNSERMANIFIRAHSEVQTVIAKII 595
QY 574 KAEINPNVVGYSFTMBEIKKAFIENKOLLATEKKYTHYNTALSVALLLTNKSSVPRVY 633
Db 596 KAIINPKTDGLTFLDELKQAFKIYNEDMRQAKKYQTSNIPTAYALMLSNKSITRLY 655
QY 634 GMFTDDGQYMAHKIYINYEAITLLKARIKYVSGGQMRNQOVGNSE-----ITS 684
Db 656 GDMYSDDGQYMAKSPYDAIDTLLKARIKYAAGGQDMKITVYVEGDKSHMDWDYTGVLTS 715
QY 685 VRYGKALKATDGTDRRTTSQVAVLEGNNPSLRKASDRVVNNCAAHKQAYPLLLT 744
Db 716 VRYGTANEATDQGEAKTQGMVITSNNPSKLQNDKVIYNMGAHKNQEYRPLLLT 775
QY 745 TNGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKGYANPQVSGYLVGVVPPVGAALIKMF 804
Db 776 TKDGLTSYTSDDAAKSLYRKTKNDKGLVDFDASDIQGYLNPQVSGYLVVWPVGAS--DNQ 833
QY 805 ALRLARPHQOMAS---VHQNAALDSRVMPGFSNFQAFATKKEEYTNVVIKQNDKFAEW 861
Db 834 DVRVAASKANATQGVYESSSALDSQLIYEGFSNFQDFVTKDSQDYTNKKIAQNVQLFSW 893
QY 862 GVTDFEMAPQVYSYSDGFLDSVIONGYAFTDRYDLGISKPNKYGTADLVKAIKALHSK 921
Db 894 GVTSEMAPQVYSYSDGFLDSIIQNGYAFBDRYDLAMSNNKNGYSGQDMINAVKALHS 953
QY 922 GIKVMADVPQMYAFPEKEVVVTRVDKYGTPVAGSQIKNTLYVVDGKSGKQOQAKY 981
Db 954 GQIVADVPQIYNLPQKEVVVTRVNDYGEYRKDSEIKNTLYAANTKSNKDYQAKY 1013
QY 982 GAFLELOAKYPELPARKQISTGVPMWDSVKIKQMSAKYFNGTNIILRGAGYVLKQDATN 1041
Db 1014 GAFLELAAPYSIENRTQISNGKIDIPSEKITAWKAKYFNGTNIILRGVGYVLKDNASD 1073
QY 1042 TYENISDNKEINFLPKTLINQDSQVGSFSDGKYVYYSYTSQVQAKNTFISEG-DKWWYFD 1100
Db 1074 KYFELKNGQ--TYLPQMTNKEASTGFVNDGNGMTFYTSYTSQVQAKNFPVQAKGNWYFD 1131
QY 1101 NNGYVMTQAQSIINGVNYVYFLSNGQLRDAILKNEIDGTVAAYGNDGRRYENGYYQF-MSGV 1159
Db 1132 NNGHMYVGLQNGEVOYFLSNGVQLRESFLENADGSKNYFQHLGNRYSGNYYSFNDISK 1191
QY 1160 WRHEN-NGEMSGLTVIDQVOYFDEMGYQAKGKFTTADGKIRYFDPKQSGNMVNRPIE 1218
Db 1192 WRYFDASGMAVGLKTINGNTQYFPDQGYQVKGAWITGSDGKRYFDDGSGNMAVNRPAN 1251
QY 1219 NEEGKWLVLGEDGAATVSGTINGHLYFRANGVQVKGFEFTDHHGRISYVDGNSGDQIR 1278
Db 1252 DKNGDWYLLNSDGTALVGVQTINGKTYFYFGDQKQIKGIITD-NGKLKYLANSSEGLAR 1310
```

```
QY 1279 NRVFVNAQOMFYFDNNGYAVTGARTINGQLLYFRANGVQVKGFEFTDRYGRISYYDGS 1338
Db 1311 NIFATDSQNNWYFSGDGVAVTGSQTIAGKLYFASDGQVKGSGFTYTN-GKVHYHHADS 1369
QY 1339 GQOIRNRFVRNAQOMFYFDNNGYAVTGARTINGHLYFRANGVQVKGFEFTDRHGRISY 1398
Db 1370 GELQVNRFEADKGNWYLLDSNGEALTGSORINDORVFFTRGKQVKGVDVYDERLLVY 1429

RESULT 12
US-10-383-930-38
; Sequence 38, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 38
; LENGTH: 1554
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-383-930-38
```

```
Query Match 44.4%; Score 3434.5; DB 4; Length 1554;
Best Local Similarity 46.6%; Pred. No. 1.4e-189;
Matches 731; Conservative 248; Mismatches 458; Indels 131; Gaps 39;

QY 1 MDKVRVYKLRKYKRWVTVSVASAVMTLLTTLSSGLVKAD---SNESKSO-----46
Db 1 MEKLLHYKLYKHVKGHWVIAVAS-IGLVSLVGAGTVAEDKVANDTTAQTAVGVDVTGQQ 59

QY 47 -ISDNTSVV---TANESNVITEATSKQEAASQTNHTVTSSSTSVVNPKEVVSNP 102
Db 60 ATTNDANTTTTDTDTADQSAN-----TNQDQAGSDQSNQDQAKQDTAN-----103

QY 103 YTVGETASNGEKLQNTT-----TVDKTS-----EAAANNISKQTTEADTVLDDNSAANL 153
Db 104 -TDRNQADNSQTDNNQATDQATSPATDGTSVQRRDAANVATAADQEGQTAPSEKSAAL 162

QY 154 QILEKLPVVKETIDGKYVYYDNNKGVRTNFTLIADGKILHFD-ETGAYTDTSDTVNKOIV 212
Db 163 ---SLDNVKLIDGKYVYVQADGSKYKKNFAITVNGOMLYFSDTGCALSTSTYSFSGQ-T 217

QY 213 TTRSNLYKKNQVYDRSAQSFEHVDHYLTASWYRPKYILKDGKTKWTOSTEKDFPLMT 272
Db 218 TNLVDDFSSHNAKYADVSTAKSFELVNGYLTANSWYRPAGILRNGQTWEASNENDLRPVLMS 277

QY 273 WWPQOETQORVYVYVNAOLGINKT-YDDTSNQLQNLIAAATQAKIEAKITTLKNTDMLR 331
Db 278 WWPQKDTQVAVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 337

QY 332 QTISAFVKTQSAWNSDSEKPFDD---DHLQNGAVLYDNEGLTPYANSNYRIILNRTPTNQTG 389
Db 338 TAMEAFVAAQPKNNMSTEN-FNKGDHLOGGALLYTN-SDLTPWANSYRLNRTPTQDQ 395

QY 390 KQDPRYTADNTIGGYEFLLANDVNSNPVQAEQLNLHFLMNFNIYANDPDANFDSIR 449
Db 396 TK--KYFTEGEGGYEFLLSNDVDNSNPVQAEQLNLHFLMNGDIIVMGDKDANFDGVR 453

QY 450 VDAVDNVADALQIOTAGDYLLKAAKGHIKNDKAAANDHLSLILEAWSNDPTLYLHDDGDNMIN 509
Db 454 VDAVDNVADLLQVYSNYFKNYKVTDSSEANALAHISILEAWSLNDNQNTDNGTALSI 513
```



```
Db 514 DNSSRLLTSIAVLTQPGQRIIDLSNISESVNKRANDTAIGDTIPTYSFVRADHSEVQTV 573
Qy 569 IADIIKAEINPNVVGYSFTMEIEIKAFIYNKDLATEKKYTHYNTALSVALLLTNKSV 628
Db 574 IAKIVKEKIDTNSDGYTFLDQLKDAFKIYNEMAKVNTYTHYNIIPAAVALLLNWSV 633
Qy 629 PRVYGDMPDQGMMAHKTINYEAIETLLKARIKIVSGGQAMRNQOV--GNSIITSVRY 687
Db 634 PRVYGDLYTDDQYMAKSPYDAIATMLQGRIAVSGGSEEVHKVNGNQNILLSSVRY 693
Qy 688 GGALKATDT--GDRTRTSGVAVIEGNNPSLRLKASDRVVVNMGAHKNQAYRPLLLTID 746
Db 694 GDLMSADDTQGTDSLRTSLGVTLSNDPNDL--CGDSLTVNMGRAHQAQYRPLILGFK 752
Qy 747 NGIKAYHSDQEAAGLVRYTNDRGELIFTAAADIKGVANPOVSGYLGVVVPVGAALIKFAL 806
Db 753 DGVQSYLKDS--TNIVKYTDANGNLFTTADDIKGYSTVDMSGYLAVWVPVGA--DGDQV 809
Qy 807 RLARPHQQA-----SVHQNALDSRVNFEFSNFQAPATKEBYTNVIAKNVDKFAEMGV 863
Db 810 RVAADTNQKADGKSLKTSALDSQIVIEGFSNFQDPANNADYTNKIAENADFFKGLGI 869
Qy 864 TDPENAPQVSSDGSFLDSVLONGYAFTRDYLGISPKNKYGTADDLVKAIKALHSKI 923
Db 870 TSFENAPQVSSATDGSFLDSIIQNGYAFSDRYDLAMSKNNKYGSKDDLALAKALHANGI 929
Qy 924 KMAWDVDPQMAFFPEKEVWVATRVYDKGTPVAGSOIKNTLVVDPGKSGKDDQAKYGA 983
Db 930 QAIADWVDPQIYQLPGSEVVTAKRTNSYGNPTFDAYINNALYATNTKSGSDYQAYGA 989
Qy 984 FLEELQAKYPELPARKQISTGVMPDPSVKIKOWSAKYFNGTNILGRGAGYVYLKQATNTY 1043
Db 990 FLEDELKAKYPMFTVMISTGKPIDPSTKIKQWEAKYFNGTNVILGKAGYVYLSDDATGY 1049
Qy 1044 FNISDNKEINFELPKLL--NQDSQVGSFDGKGYVYVYSTSGVQAKNTPISBGDKWYFNN 1102
Db 1050 FTVNENG--FLPASFTGDQNAKTFYDGTGMAYIYSTSGNKAIVSYEGGHYFYFDK 1107
Qy 1103 GMYVTGA--QSLNGVNYFLSGLQLRDAILKNEGTGYAYGNDGRYE--NGYQFM---- 1156
Db 1108 GMYVTGSYKAEQNDYFLPNGIQWRDALYDAQGNSYYGRTGLYKGNWYFVDPNN 1167
Qy 1157 --SGVWRHFN--NGEMSVGLTVIDGQVQYFDEMGOYQAKGFVTTADGKIRYFDKQSGNMYR 1213
Db 1168 ANKTVRYFDANNVMAIGVRNMYGTYFYFDENGFOAKGQLLTDDKG--THYFDEDNAMAK 1226
Qy 1214 NRTFENESGKWLGLGEDGAATGSGTINGQHLYFR--ANGVQVKGFEVTDHHRISYDGN 1272
Db 1227 NKV--NVGDDWYMDGNGNAVKGOYPVNNQILYFNPETGVQVKGQFIIDAQGRTSYDAN 1285
Qy 1273 SGQDIRNRF-----VRNAQOGWFYFD--NNGYAVVTGARTINGQL 1309
Db 1286 SGALKSSGFFTPNGSDWYTAENGYVYKGFQVAENQDQWYFDTQTGQAKGAQVGDGRD 1345
Qy 1310 LYFRA--NGVQVKGFEVTDYRGRISYDGNSGDQIRNRVRNAQOGWFYFDNNGYAV---- 1364
Db 1346 LYFNPDSGVQVKGDFATDESNTSFYHGDNGDKVVGFFTTGNNAWYADNNGNLVKGFG 1405
Qy 1365 -----TGARTINGQHLYFRA--NGVQVKGFEVTDYRGRISYDGNSGDQ 1406
Db 1406 BIDGKWHFDEVGTQQAAGALVNGQYLYFDVDSGIQVKGDFVTDGQNTSYVDVNSGDK 1465
Qy 1407 IRNRFVRNAQOGWFYFDNNGYAVVTGARTINGQHLYF--PANGVQVKGFEVTDYRGRISYD 1465
Db 1466 KVGFFTTGDNWYADQGNLAKGRKSIDNQDLYFDPATGQVKGQLVSD--GRNYFYD 1524
Qy 1466 ANSGERV 1473
Db 1525 SGSGNMAK 1532
```

RESULT 14

US-10-383-930-40

```
; Sequence 40, Application US/10383930
; Publication NO. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 40
; LENGTH: 1518
; TYPE: PRT
; ORGANISM: Streptococcus salivarius
; US-10-383-930-40
```

Query Match 41.3%; Score 3196.5; DB 4; Length 1518;
Best Local Similarity 45.2%; Pred. No. 8e-176;
Matches 695; Conservative 233; Mismatches 472; Indels 139; Gaps 31;

```
Qy 1 MDKVRVYKLRKVKRWTVSVASAVMTLTTLTSSGL-----VKADSNESKSOISNDSNTS 54
Db 1 MENIKHYKLHKVKQWTVIAVAS--VALATVLGGLSVTSSVSADETQKVTYQNSGTT 58
Qy 55 -----VVTAENEENSVITEATSKQEAASSQTNHVTVTSSSSTS 91
Db 59 ASLVTSPBATKEADKRTNTKEADVLTPAKETNAVETATTTNTQATAEAATTATTADVA 118
Qy 92 VVNPKEVSVNPTVGETASNGEKLNQNTTVDKTSAAANNISKOTTTEADTDVIDDSNA 151
Db 119 AVPNKEAV---VTTDAPAVTTEKABEQPATV---KAEVVNTEVKAPEA---ALKDS--- 165
Qy 152 NLQILLEKLPNVEIDGKYVYVNNNGKVRNTFTLIADGKILHFDGTGAVTDTSIDTVNKDI 211
Db 166 EVEAALSANKNTKIDGKYVYVNNEDGSHKENFAITVNGQLYFGKGCALITSSSTYFTPG- 224
Qy 212 VTTRSNLYKYNQVYDRSAQSFHVHDYLYLTAEWSYRKYILKDGKTWTQSTEKOPRPLM 271
Db 225 TTNVVDGFSINNAYDSEASFELIDGYLTADSWYRPASIIKDGVTWQASTAEDFRPLM 284
Qy 272 TWPDQETORQYVNMNAQLGINKTYDDTSNQLNLNIAAATIOAKIEAKITTLKTDMLR 331
Db 285 AWWPNVDQVNYLNTMSKVFNLDAKYSSTDQETLKVAAKDIOIKIEQIOAKESQWL 344
Qy 332 QTISAFVKVTQSAWNSDSEK---PFDDHLQNGAVLYDNEGKLTVPVANSYRILNRTPTNQ 387
Db 345 ETISAFVKTPQWKNKETYENSKGGEDHLOGALLYVNDNR--TPWANSYRRLNRTATNQ 403
Qy 388 TG-----KKDPRYTADNTIGYEFLLANDVNSNPVQAEQLNWLHFLMNFNIYA 438
Db 404 TGTIDKSLDSQSDP-----NEMGDFLLANDVDLSNPVQAEQLNQIHYLMNWSIVM 458
Qy 439 NDPANFDSIRVDAVDNVDADLLQIAGDYLKAAGIHKNDKAANDHLISILSAWSNDPTPY 498
Db 459 GDKDANFDGIRVDAVDNVDADMLQLYTNYFREYGVNKSSEANALAHISVLEAWSLNDHY 518
Qy 499 LHDDGDNMINMDNKLRLSLFSLAKPLNORS--GNPLITNSLVNRTDNDNAET----- 549
Db 519 NDKTGAALAMENKQRLALLFSLAKPIKERTPAVSPLYNNTF--NTQDEKTDWINKDGS 577
Qy 550 -----AAVPSYSFIRAHDSVEQDLIADIKAEINPNVNVGYSFT 587
Db 578 KAYNEDGTVKQSTIGKYNEKYGDASGNVYFIRAHNNVODIIAEIIEIKKEINPKSGFTIT 637
Qy 588 MEEIKKAFIYNKOLLATEKKYTHYNTALSVALLLTNKSVPRVYVYGDMPDQGMMAH 647
Db 638 DAEMKQAFIYNKMLSSDKKYTLNNIPAAVAVMLQNMMETITRVVYVYGDLYTDDGHYMETK 697
```


THIS PAGE BLANK (USPTO)

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	454	5.9	2710	7	US-11-051-453-41	Sequence 41, Appl	
2	395	5.1	726	7	US-11-052-554A-217	Sequence 217, Appl	
3	382	4.9	2367	7	US-11-051-453-42	Sequence 42, Appl	
4	283.5	3.7	396	7	US-11-022-562-228	Sequence 228, Appl	
5	261.5	3.4	658	6	US-10-873-528-17	Sequence 17, Appl	
6	261.5	3.4	677	6	US-10-873-528-155	Sequence 155, Appl	
7	231	3.2	2314	7	US-11-013-759-11	Sequence 11, Appl	
8	248	3.2	619	7	US-11-052-554A-229	Sequence 229, Appl	
9	247	3.2	2902	7	US-11-052-554A-91	Sequence 91, Appl	
10	240	3.1	701	7	US-11-052-554A-231	Sequence 231, Appl	
11	235.5	3.0	1155	6	US-10-793-626-1780	Sequence 1780, Ap	
12	221.5	2.9	744	6	US-10-873-528-184	Sequence 184, Appl	
13	221	2.9	338	7	US-11-052-554A-228	Sequence 228, Appl	
14	219	2.8	1992	7	US-11-013-759-3	Sequence 3, Appli	
15	219	2.8	1992	7	US-11-013-759-13	Sequence 13, Appl	
16	219	2.8	2047	7	US-11-013-759-4	Sequence 4, Appli	
17	219	2.8	2047	7	US-11-013-759-7	Sequence 7, Appli	
18	219	2.8	2053	7	US-11-013-759-9	Sequence 9, Appli	
19	215.5	2.8	690	7	US-11-052-554A-232	Sequence 232, Appl	
20	212.5	2.7	1588	7	US-11-052-554A-280	Sequence 280, Appl	
21	211.5	2.7	2399	7	US-11-052-554A-92	Sequence 92, Appl	
22	211	2.7	483	6	US-10-630-203-10	Sequence 10, Appl	
23	211	2.7	514	7	US-11-002-188-6	Sequence 6, Appli	
24	210	2.7	483	6	US-11-020-203-8	Sequence 8, Appli	
25	210	2.7	483	6	US-10-630-203-30	Sequence 30, Appl	

QY 1264 GRISYYDNGSGDQIRNRFVRNAQGWFFYDNNGYAVTG-----ARTIN----- 1306
Db 2142 -----VQNIIDDNYFYIDONGIVQIGVFDTSQGYKYPAPANTVNDNIY 2183
QY 1307 GQLLYFRANGVQVKGFEV-----TDRY----- 1330
Db 2184 QGAVEY--SGLVRGDEVYFGETYTIETGWLYDMENESDKYYFVPETKACKGINLIDD 2241
QY 1331 ISYDNGSGDQIRNRFVRNAQGWFFYDNNGYAVTGARTINGOHLHYFRANGVQVKGFEV 1390
Db 2242 IKYFDEKG--IMRTGLISFENNYYFNEBGEIQGYINIEDKMFYGEDGVQMVGVENT 2299
QY 1391 DRHGRIYYDNGSGDQIRNRFVRNAQGW-----W-----FYDNNGYAVTGARTING 1437
Db 2300 P-----DGPKYFAHQNTLDENPEGBSINYTGMLGLDEKRYFYFTDVIATGSGVLIIDG 2351
QY 1438 QHLYFRANGVQV 1449
Db 2352 EBYFDPDTAQL 2363
RESULT 4
US-11-022-562-228
; Sequence 228, Application US/11022562
; Publication No. US20050249742A1
; GENERAL INFORMATION:
; APPLICANT: Ruprecht, Ruth M.
; APPLICANT: Shisong, Jiang
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING
; TITLE OF INVENTION: A CYTOTOXIC T LYMPHOCYTE IMMUNE RESPONSE
; FILE REFERENCE: DFN-043CN
; CURRENT APPLICATION NUMBER: US/11/022,562
; CURRENT FILING DATE: 2004-12-22
; PRIOR APPLICATION NUMBER: PCT/US03/20322
; PRIOR FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: 60/392718
; PRIOR FILING DATE: 2002-06-27
; NUMBER OF SEQ ID NOS: 340
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 228
; LENGTH: 396
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-11-022-562-228

Query Match 3.7%; Score 283.5; DB 7; Length 396;
Best Local Similarity 29.9%; Pred. No. 2.8e-09;
Matches 96; Conservative 44; Mismatches 124; Indels 57; Gaps 17;
QY 1144 DGRYVYNGYQFMGSGVWRHFN--GMS---VGLTVIDGQVYFDEMGOAKGKFTVTTADG 1199
Db 77 DIKKLSLGY--IMSN--FKSFENSELDRHLGFKIIDNKTYDYDEASKLVKG--LININN 131
QY 1200 KIRYEDKSGNMYRNFLENBEGKWLXG--BDGAATGSGTQINGOHLHYFRANGVQVKGFE 1258
Db 132 SLFYFDPDIESLVGT--WQTNGKYYFDINTGAASTSYKINGKHFYFNNGVMQLGVF 189
QY 1259 V-----TDHGRISYYDNGSGDQ---IRNRFVRNAQGWFFYDNNGYAVTGARTINGQLLY 1311
Db 190 KPGDGFYFAPANTONNIEGQAIYVQSKFL--TLNGKYYFDNDSKAVTGWTIDGKKY 248
QY 1312 FRANGVQVKGFEVTDYGRISYYDNGSGDQIRNRFVRNAQGWFFYD--NNGYAVTGARTI 1370
Db 249 FNLNTAEATGWT-----IDGKKYYFNTNTSIATGYTII 284
QY 1371 NGOHLHYFRANGVQVKGFEVTDHGRISYY-----DGNS--GDQIR--NRFVRNAQGWFFY 1421
Db 285 NGKHYFNTDGMQIGVTF--KGPNGFEYFAPANTDANNIEGQAIHYQNRFL--YLHDNIYY 341
QY 1422 FDNNGYAVTGARTINGOHLHYF 1442
Db 342 FGNSKAVTGWTINGNVYVF 362

RESULT 5
US-10-873-528-17
; Sequence 17, Application US/10873528
; Publication No. US20050276814A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129WO
; CURRENT APPLICATION NUMBER: US/10/873,528
; CURRENT FILING DATE: 2004-06-23
; PRIOR APPLICATION NUMBER: US/09/769,787
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 17
; LENGTH: 658
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-873-528-17

Query Match 3.4%; Score 261.5; DB 6; Length 658;
Best Local Similarity 20.7%; Pred. No. 1e-07;
Matches 97; Conservative 72; Mismatches 188; Indels 111; Gaps 18;
QY 1029 RGAGYVLKD--QATN-----TYFNISONKEINFLPKTLNODSQVGSYDQKGVY 1077
Db 29 QGQYKLEDGSAANEWVFDTHYQSWFYI--KADANYAENEWLKQ-----GDDYFY 77
QY 1078 YSTGQAKNTFISEGDKWYVDNNGVMVTGAQSGINGVNYFLSNGLOLRDAILKNEBGT 1137
Db 78 LKSGGYMAKSEWEDKGAFFYLDQDGKMRNAV--GTSYVGATGAKVIEDWVYDSQYDA 135
QY 1138 YAYTGNDGRRYGYQFMGSGVWRHFNNGEMSVGLTVIDGQVYQVYFDEMGOAKGKFTVTTA 1197
Db 136 WFIYKADQHAKEWLQ-----IKGDIYFKSGGYLLTSQWINGA 175
QY 1198 DKIRYFDKQSGNMYRNFLENBEGKWLXGDEGAAVTGSGQTI--NGOHLHYFRANGVQVKG 1256
Db 176 ----YVNASGAKVQQGWLFDKQYQSWFYIKENGNYADKEWIFENGHYIYLLKSGGYMAAN 230
QY 1257 EFVTDHGRISYYDNGSGDQIRNRFVRNAQGW--WFYDNNGYAVTGARTINGQL----- 1309
Db 231 EWIWDKES--WFYLLKFDGKMAEKWVYDSHSAQWYIFKSGGY-----MTANEWIWDKESW 283
QY 1310 LYFRANGVQVKGFEVTDYGRISYYDNGSGDQIRNRFVRNAQGWFFYDNNGYAVTGART 1369
Db 284 FYLKSDDGKIAEKWVYDSHSAQWYIFKSGGYMTANEWIWDKE--SWF----- 328
QY 1370 INGOHLHYFRANGVQVKGFEVTDHGRISYYDNGSGDQIRNRFVRNAQGW-----GQWYFDDN 1425
Db 329 ----YLLSDGKIAEKWVYDSHSAQWYIFKSGGYMAKNETVDGYQLGSDGKWL----- 377
QY 1426 GYAVTGARTINGOHLHYF-----ANGVQVKGFEVTDYGRISYYDANS 1468
Db 378 ----GGKTTNENAAVQVVPVANTVYDSDGE-----KUSIYISQGS 413

RESULT 6
US-10-873-528-155
; Sequence 155, Application US/10873528
; Publication No. US20050276814A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M

```
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P211290
; CURRENT APPLICATION NUMBER: US/10/873,528
; CURRENT FILING DATE: 2004-06-23
; PRIOR APPLICATION NUMBER: US/09/769,787
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 155
; LENGTH: 677
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-873-528-155

Query Match
3.4%; Score 261.5; DB 6; Length 677;
Best Local Similarity 20.7%; Pred. No. 1.1e-07;
Matches 97; Conservative 72; Mismatches 188; Indels 111; Gaps 18;

QY 1029 RGAGVYVKD--QATN-----TVFNISDNKEINFLPKTLNODSQVGFSDGKGVY 1077
Db 48 QKQYLKEDGSQANEFVDTHTYQSWFYI--KADANYAENELKQ-----GDDYFY 96

QY 1078 YSTSGYQAKNTFISEGDKWYFDNNGYMTVGAQINGVNYFLSNGLQLRLDAILKNEGT 1137
Db 97 LKSGGYMAKSEWEDKGFYLLDQDGKMKRANW--GTSYVGATGAKVIEDWYDSQYDA 154

QY 1138 YAYIGNDGRYNGYQFMGVRHFNNGEMSVGLTVIDGVOYQFDEMGOAKGPFVTTA 1197
Db 155 WFYIKADGQHAKEWLIQ-----IKGXDYFKSGGGLLTQWQINQA 194

QY 1198 DGKIRYFDKQSGWNYRNRIENECKWLYLGEDGAVTGSGQTI--NGCHLYFRANGVQVG 1256
Db 195 -----YVNSAGAKVQOGWLFQYQSWFYKENGNYADKEWTFENGHYLYLKSGGYMAAN 249

QY 1257 BEVTDHGRISYDNGSGDQIRNRFVRNAQGO--WFYFDNNGYAVTGARTINGOL----- 1309
Db 250 EWIWDKES--WFLKFDGMAKEWYDSSHQAWYFKSGGY-----MTANEWIWDKESW 302

QY 1310 LYFRANGVQVKGFEVTDVGRISYDNGSGDQIRNRFVRNAQGOQWYFDNNGYAVTGART 1369
Db 303 FYLKSPGKIAKEWYDSSHQAWYFKSGGYMTANEWIWDKE--SWP----- 347

QY 1370 INGCHLYFRANGVQVKGFEVTDVGRISYDNGSGDQIRNRFVRNAQ-----QWYFDNN 1425
Db 348 -----YLKSDGKIAKEWYDSSHQAWYFKSGGYMAKNETVDGYQLGSDGKWL----- 396

QY 1426 GYAVTGARTINGOHLFYR-----ANGVQVKGFEVTDVGRISYDANS 1468
Db 397 -----GGKTENEAAYQVVPYANVYDSGE-----KLSYISQGS 432

RESULT 7
US-11-013-759-11
; Sequence 11, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; FILE REFERENCE: 1038-921MIS:jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 11
; LENGTH: 2314
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-11

Query Match
3.2%; Score 251; DB 7; Length 2314;
Best Local Similarity 19.4%; Pred. No. 2.4e-06;
Matches 346; Conservative 182; Mismatches 655; Indels 596; Gaps 79;

QY 38 ADSNESKSOISNDS-----NTSVVTANEESNVITEATSKQEAAS-----SQTNHTVTTTS 86
Db 714 ASGNDTKNIIRGLSPPLPSITNAGGVRTTEQGNITISDEDEKSKAAASIGILNFGNLKNN 773

QY 87 SSSTSVNPKEVVSNPYTVGETASNGEKLNQNTTV-----DKTSEAAANNISKQTTE 139
Db 774 SNSVGFVSTYNTVDFIDGNATTAKVYDETQNTSKVYDVNVDEKTIETLTDGNGKTNKIG 833

QY 140 AUTDVIDDSNAANLQILEKLPNVKEIDGKYVYDNNGVKRTNFTLIADGKILHFDGTGAY 199
Db 834 VKTTTLTTTNA-----NGKA--TNFSTTDNDALVNAKDIA-- 866

QY 200 TDTSIDTVNKDITVTRSNLYKKYNQVYDRSAQSFEHVDHYLTAESWYRKYILKDKGTWT 259
Db 867 --ENLNTLAKEIHTTKGTA-----DTALOTFK-----VKDQATDD 900

QY 260 QSTEDKDFRLLMTWPFDOETQRYVNYM-----NAQLGINKTY 297
Db 901 ET-----ITVGKDGTONKTVNTLKLKSGELTVATNKDGTVPFGINTQSGLKGAD 951

QY 298 DTSNQLQNLIAAATIQAKIE-----AKITTLQNTDMLRQTISAFVKTQSAMNS--- 346
Db 952 STTLNKGDLGSLKNPASNEQIQVGADGVKPAKVGKNSGTIGDT--SRITKQDQIGFTGANG 1010

QY 347 --DSKPP--DDHLQNGAVLYDNEG-----KLTPY-----ANSYRLN----- 381
Db 1011 SLDTTKPHLTQDKLKVGEVEITNTGINAGKKITNIQSGDITQNSNDVATGGRVYDLKTE 1070

QY 382 -RTPNTQCKKDPRTADNTIGGYEFLANDVDSNPVQAEQLNWLHFLMFGNIYAND 440
Db 1071 LESKINSAAK-----TAQNSL--HEFSVADEQGN-----HFTV-----SN 1103

QY 441 PDANFDSIRVDVNDVADLQIAGD---YLKAAKGIHK--NDKAANDHLSILEAWSND 495
Db 1104 PYSSYDTGKT-----SDVITFAGENGITTKVNGKVRVGVGIDQTKGLTTPKLTGVNNGG 1156

QY 496 TPLYLD--DGDNMWMDNKLRLSLFLSLAKPLNQSGMPLITNSLVNPTDN--AETAA 551
Db 1157 KGIVIDSKDQNTIT-----GLSNTLANVTN--DGAGHALSQGLANDTDKTRAASIGD 1207

QY 552 VPSYFIRAHIDSEVODLIA-----DIKAEINPNVVGYSFTMEEIKKAFIY--NKDLLA 604
Db 1208 VLNAGFNLOQGEAVDFVSTYDVFIDGNATTAKVYDDTSKTSKVYDVNVNDKNTIEV 1267

QY 605 TEKK-----YTHYNT----- 614
Db 1268 TSDKGLGVKTTTLTKTSANGNATKFSAADGDALVKASDIATHLNTLAGDIQTAKGASQAS 1327

QY 615 -ALSVALLLTNK---SSVPRVYVGMFTDDGGYMAHKTIYNEAI-----ETLLKARI 662
Db 1328 SSASYVDADGNKVIYDSTDKKY--QVNDKGVQDNKKEVAKDKLVAQAQTPDGTTLAQMNV 1385

QY 663 KYVSGQAMRNQOVGNSEIITSVRYGKALKATDGTTRTTSVAV----- 709
Db 1386 KSV-----INKEQVNDANKQKGINEEDNAFIKGLENAAKDTKTKNAAVTVGDLNNAVATPL 1440

QY 710 -IEGNNPSLRKASDRVVVNMGAHKNQAYRPLLLTTDNGIKAYHSDQBAAGLVRYT--- 765
Db 1441 TPAQDTGTTAKKLGGETLTIKGGQTDNK-----LTDNNIGV-----VAGTDGFTVKL 1487

QY 766 -----NDRGELIIFTAADIKGYANPQVSGYLGWVPVGAALIKMPALRLARPHQOM 815
Db 1488 AKDLTNLNSVNAGG---TRIDEKGISFVDANGQAKANTPVLISA---NGLDLAGKRIISNIG 1541
```


Qy	816	ASVHQNALDSRVMEFSGSNFQAPATKKEBYTNVVIKNDVKFAEWGVTDFEMAPQYVSS	875
Db	1542	AAVDDNDAV-----NFQNEVAKTVNNLNNQSGASLPFFVVTDANGKP--ING	1589
Qy	876	TDSFSLDSVIQNGYAFTDRYDLGISKENKYGTADDLKAJKAHLSKGIKVMADWVPDQMY	935
Db	1590	TDG-----KPO-----KAIKGADGKYHHANANGVPVDKD	1618
Qy	936	AFPEKEVVTATRVDDKYGTPV-AGSQ-----IKNTLVVVDGKSSGKQQA	978
Db	1619	GKPIITDADKLANLAHCKPLDAGHQVVASLGGNSDAITLTNIKSTLPQIDTPWTG--NA	1675
Qy	979	KYGAFLLELOAKYPELPARKQIS-----T	1003
Db	1676	NAGQA-----QSLPSLSAAQQSNAASVKDVLNFGNLQTNHNQVDFVKAYDTVNFVNGT	1729
Qy	1004	GVPM-----DPSVKIKQWSAKYFNGTNTILGRG---AGVVLK	1036
Db	1730	GADITSVRSDGTMSNTIVNTALAAATDDGNVLKAKDGRFYKADDDLMPNGSLKAGKAS	1789
Qy	1037	DOATNTYFNI-----SDNKEINFLPKTLNQD-----SQVGSFYDGK--GYV	1076
Db	1790	DAKTPTGLSLVNPWAGKSGTGAVALNLSKAVPKSKDGTNTTTSVSDGISIQCKDNSSI	1849
Qy	1077	YYSYSGYQAKNTFISEGDKWYFPDNGVMYMTGAOSINGVNYVYFLSGLQLRDAILKNEGD	1136
Db	1850	TLSKDGLNVGKVVISNVGK-----GTKDITDAANVOQLNE--VRNLLGLGNAGNDNADG	1900
Qy	1137	TYAYV-----GNDGRRYENGYYQF-----MS	1157
Db	1901	NQVNIADIKKDPNSGSSSNRTVIKAGTVLGKGNNDTEKLATGQVGVGDKNANGDLS	1960
Qy	1158	GVW-RHFNNGEMSVGLTVID--GOVQVFDENGVOAKGKFVTTADGKIRYFDKQSGNMY--	1212
Db	1961	NVMVKTKDGSKKALLATYNAAQGTNLTNNPAAIDRI---NEQGRFHFVHDNGQEPV	2017
Qy	1213	---RNRFIENEKGW-LYLG-----EDG--AAVTGSQTTINGOHLYFRANGVQVKE---	1257
Db	2018	VQGRNGIDSSAGSKHSVAIGFOAKADGEAAVAIGRTQAGNQSIAGIDNAQATGDSQIAI	2077
Qy	1258	---FVT-DHHGRI-----SYDNGSGDQIRNRFVRNAQGWQFYFDNN-----	1295
Db	2078	GTGNVVTGKHSIGAIGDPSVKADNSYSVGN-----NOFIDATQDVFVGNNITVTESN	2132
Qy	1296	-----GYAVTGARTINGQLLYFRANGV--QVKGFEVTDRYGRI SYDNGSGD	1340
Db	2133	SVALGNSAISAGTHAGTQAKSGDGTAGTTTTAGATGTGVPFAGQTAVGAVSV--GASGA	2190
Qy	1341	QIRNRFVRN-AQOGWIFYDNNGYAVTGARTINGOHLYFRANGVQVKEFVTDHRGRISYY	1399
Db	2191	E---RR'ONVAAGE-----VSATSTDAVNGSOLYKA'QGI---ANATNELDHR'IQON	2236
Qy	1400	DGNSGDOIRNRFVRNAQGWQFYFDNNGYAVTGARTINGQ	1438
Db	2237	ENKANAGISSAMAMASMPQ-AYIPGRSMVTCGGTATHNGQ	2274

RESULT 8

```

RES001 6
US-11-052-554A-229
; Sequence 229, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589, 227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06

```

```

; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 229
; LENGTH: 619
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae R6
US-11-052-554A-229

Query Match      3.2%; Score 248; DB 7; Length 619;
Best Local Similarity 24.3%; Pred. No. 5.5e-07;
Matches 84; Conservative 35; Mismatches 129; Indels 98; Gaps 13;

Qy      1056 P K T L L N D S Q V G S Y D G K G V Y Y S T S G Y -----A K N T F I S E G D K W Y Y F D N N G Y 1104
Db      360 P K P E K T D D Q A E E D Y A R R S B E E Y N R L T Q Q P P K A E K P A P A P K T G K W Q E G M W Y F Y N T D G S 419

Qy      1105 M V T G A O S I N G V Y Y F L S N G L Q L R D A I L K N E D G T V A Y Y G N D G R R Y E N G Y Y Q P M S G V W R H F N 1164
Db      420 M A T G W L Q N N G S W Y Y L N S N G A M -----A T G W L Q Y - N G S W Y Y L N 455

Qy      1165 - N G E M S V G L T V I D G O V Y F D E M G Y O A K G F V T T A D G K I R Y F D K O S G N Y R N R F I E N E S G K 1223
Db      456 A N G A M A T G W A K V N G S W Y Y L N A N G -----A M A T G W L Q Y -----N G S 490

Qy      1224 W L Y I G E D G A A V T G S O T I N G Q H L Y F R A N G V O V K G E F V T D H G R I S Y I D G N S G D Q I R N R F V R 1283
Db      491 W Y Y L N A N G A M A T G W A K V N G S W Y Y L N A N G A M A T -----G W L Q Y -----527

Qy      1284 N A O G W F Y F D N N G Y A V T G A R T I N G Q L L Y F R A N G V O V K G E F V T D R Y G R I S Y Y D G N S G D Q I R 1343
Db      528 --N G S W Y Y L N A N G A M A T G W A K V N G S W Y Y L N A N G A M A T G -W V K D --G D T W Y Y L E A S G A M K A 582

Qy      1344 N R F V R N A Q O G W F Y F D N N G Y A V T G A R T I N G Q H L Y F R A N G V O V K G E F V 1389
Db      583 S O W F K -V S D K W Y Y N G L G -A L A V N T T V D G Y K -----V N A N G E W V 619

```

RESULT 9

```

US-11-052-554A-91
; Sequence 91, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 91
; LENGTH: 2902
; TYPE: PRT
; ORGANISM: Helicobacter pylori J99
US-11-052-554A-91

```


Db 1010 IDSGFYTKPKYSVDYVWEDTNK-----DG-IODNEKGISGVKVTLKDEKGNII 1058
QY 1195 TTA-----DGKIRYFDKQSNMYRNRRIENEEGKWLVLGEGDGAATGVSQTIHQHLYFRAN 1250
Db 1059 STYTIDENKQYQDNLDGN-YIHP-EKPEG-----MTOTTANSNGNDDKAD 1105
QY 1251 GVQVKGEFVTDH-----HGRISYYDGN 1273
Db 1106 GEDVR-VTITDHDDFSIDNG---YFDDDS 1130

RESULT 12

US-10-873-528-184
; Sequence 184, Application US/10873528
; Publication No. US20050276814A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hanbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129WO
; CURRENT APPLICATION NUMBER: US/10/873,528
; PRIOR FILING DATE: 2004-06-23
; PRIOR APPLICATION NUMBER: US/09/769,787
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 184
; LENGTH: 744
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-873-528-184

Query Match 2.9%; Score 221.5; DB 6; Length 744;
Best Local Similarity 29.1%; Pred. No. 2.3e-05;
Matches 69; Conservative 29; Mismatches 118; Indels 21; Gaps 10;
QY 1218 ENEEGKWLGBDGAATGVSQTIHQHLYFRANGVQVKGFEVTDHGRISYYDGNSGDQI 1277
Db 529 KGENGWYFYNTDGSMAICWLQNGSWYLLNANGAMATG-WVKD--GDTWYLEASGAMK 585
QY 1278 RNRFRNAGQWFYDNNGYATGARTINGQLLYFRANGVQVKGFEVTDHGRISYYDGN 1337
Db 586 ASQWFK-VSDKYYVNSNGAMATGMLQYNGSWYLLNANGMATGWL--QYNGSWYLLNA 641
QY 1338 SGDOIIRNFRVNAQGWFEYDNNGYAVTGARTINGQHLYFRANGVQVKGFEVTDHGRIS 1397
Db 642 NGDMATG--WAKVNGSWYLLNANGAMATGWAKVNGSWYLLNANGSWATG-WVKD--GDTW 696
QY 1398 YYDGNSGDOIIRNFRVNAQGWFEYDNNGYAVTGARTINGQHLYFRANGVQVKGFEV 1454
Db 697 YLEASGAMKASQWFK-VSDKYYVNYGLG-ALAVNTVDGK-----VNANGWV 744

RESULT 13

US-11-052-554A-228
; Sequence 228, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004

; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 228
; LENGTH: 338
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae R6
US-11-052-554A-228

Query Match 2.9%; Score 221; DB 7; Length 338;
Best Local Similarity 21.9%; Pred. No. 8.2e-06;
Matches 84; Conservative 59; Mismatches 130; Indels 110; Gaps 18;
QY 1117 YYFLSNGQLRDAILKNGDGYVYVYGNDRRRYENGYYQFMSGVWRHFNNGENSVGLTVTD 1176
Db 15 FFF---GLLATNTVANTTGG-RFVDKDKRKY-----YVKDDHKAIIYWHKID 57
QY 1177 GQVQYFDEMGYQAKGFVTTADGKIRYFDKQSGNNYRNRFIENE-----EGKWLYLE 1229
Db 58 GKTYFYFGDIGEMVVG-----WQYLE-IPGTGYRDLFDNQPVNEIGLQEKWYFQG 107
QY 1230 DGAHV--TGSQTING-----OHLIFRANGVQVKGFEVTDHGRISY 1268
Db 108 DGALLEQTDKQVLEAKTSBNTGKYGEQYPLSAEKRTTYFDNNYAVKTGWIYE-DGNWY 166
QY 1269 YD--GNSGDOIIRNRFV-----RNAQGWFEYDNNGYAVTGARTING 1307
Db 167 LNKLGNGFDDSDYNPLPIGEVAKGTQDFHTVIDIRSKPAPWYLLDASGNLTDWQKVG 226
QY 1308 QLLYFRANGVQVKGFEVTDHGRISYYDGNSGDOIIRNFRVNAQGWFEYDNN-NGYAVTG 1366
Db 227 KWYFYFGSGSMATG-----XKVR--GKWWYLDNKGDMKTG 261
QY 1367 ARTINGQHLYFRANGVQVKGFEVTDHGRISY--DNGSDQIRNFRVNAQGWFEYDNN 1425
Db 262 WQYLGKWKYLLRSSGAMVTGWY---QDGLTWYLLNAGNMDKMTGWFOVN--GKWWYAYS 316
QY 1426 GYAVTGARTINGQHLYFRANGVQ 1448
Db 317 G-ALAVNTVDGYSVNYNGEWVQ 338

RESULT 14

US-11-013-759-3
; Sequence 3, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; FILE REFERENCE: 1038-921MIS:Jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1992
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-3

Query Match 2.8%; Score 219; DB 7; Length 1992;
Best Local Similarity 20.0%; Pred. No. 0.00013;
Matches 351; Conservative 215; Mismatches 665; Indels 524; Gaps 93;
QY 30 TLGGGLVKADS--NESKQISQNSDNTSVVTANEESNVI-----TEATSK 71
Db 119 TLINDLINGHVLKEIRSSKNDVKYRRTTASGHASTAVGAMSYAQGHFSNAGFGRATATK 178

QY 72 QE-----AASSQNHVTTSSST-----SVNPKEV 98
DB 179 SAYSLAVGLAATAEQSTIASGDSATSSGLAIALGAGTRAQLOGSIALQGSVVQTSDN 238
QY 99 VSNPVTGETASGKLGQNTTVDKTSSEAAANNISKO--TTEADTVDDSDNAANLOIL 156
DB 239 NSRPAYTPNTQALDPKFAQNTWKAGPLISIGNSIKRKLIIVGAGVKNKTDVAVNAQLEAV 298
QY 157 EKLPNVKEIDGKYVYDNG-----KVRTNFTLIADGKILHFDGTAYTSDTIVNK-DIV 212
DB 299 VKWAKERRIT--FQDDNDSTDKGLDNLTKGGA---ETNALTDNNIGVVKREADNS 351
QY 213 TTRSILYKYNQVDRSAQSPHVDHLYTASWYRPKYLKDKTWTQSTEDKFRPLMT 272
DB 352 GLUKVLAKTLNLTNTVNTLNAITTVKVGSSSTTAELLSDSLTFTQFNT-----402
QY 273 WMPDQETORQY-----VNYMN-----AOLGINKTYDD-----TSNQL 304
DB 403 --GSQSTKTVGVNGVFTNNAETTAALITTRIDRKIGFARDGDVDEKQAPYLDKKQL 460
QY 305 QLNIAAATIOAKIEAKITTLKNVDLROTISAFVKTQSAWNS---DSEKPFDDHLQNGAV 361
DB 461 KVGSAIITIDNGIDAG-----NKKISNLAKGSSANDAVTIEQLKAAKPTLNAGAG 510
QY 362 LYDNEGKLTPIYANSYRILNRT-PTNQTQKDPRTADNTIGGYEFLANDVDSNPVVQ 420
DB 511 ISVPTFISVDKSG---NVTAPTNYIGVKTTELSNDSGTSKFS---VKGSGTNSLVT 563
QY 421 AEOL-NWLH-----FLMNFNGIYANDPDANFDSIRVDAVDNVA-DLLQIAG-----465
DB 564 AEHLASYLNEVNRADALSQFTVKEEDDDANAITVAKDTYKNAGAVSILKKGKNGLT 623
QY 466 -----DYLKAAKGIRKND-----KAANDHLSILEAWSNDTPLYLHDDGN 505
DB 624 VATKKGTVTFGLSQDGLTIGKSTLNDGLTVKDTNEQIQV---GANGIKFTVNGSN 679
QY 506 -----MINMDNKLRLSLS-----LAKPLNQRSGM---NPLIINSLVNRDNDNAETAA 551
DB 680 PGTGIANTARITRDKIGFAGSDGAVDTNKPYLQDKLQGVNKKITNTGIN-AGGRAITGL 738
QY 552 VPSYFIRAHDS---EVODLADIITK---AEINPNVGVSTFMBEIKKAFE-IYNKDLL- 603
DB 739 SPTLSIADQSRNIELGNTQDKDKSNAASIN-DILNTGFLKNRNNPIDFVSTYDIDV 797
QY 604 -----ATEKKYTHYNALSAYALLTNKSPRVVYGDMTDDGOYMAHKTINYEAIETLL 658
DB 798 FANGNATTATVTH-DTA-----NKTS--KVVI-DVNVDD-----TTIHLTGTDNKK 839
QY 659 KARIKYV-----SGQAMRNOQVGNSE-----IITSVRYGKGALKATD 696
DB 840 KLGVRTTKLNTKSANGTATPNVNSSDEDALVNAKOIAENLNTLAKSIHTTKG-----893
QY 697 TGDTRTSGVAVIBGNPNPSRLKASDRVVMGAAHKNQAVRPLLLTTDNGIKAYHSDQ 756
DB 894 TADTALQFTVKKVDENN-----NADDANAITVGQKNANNQVNTLTLKGENGNI-----943
QY 757 EAAGLVRYTNRGELIF---TAADIKGYANPOVSGYLGWVVPVGAALIKMFALRLARPHQ 813
DB 944 -----KTDKNGTGVFGINTTSLGKAGKSTLNDGGLSIKNPTGSEQIQVGA-----DGV 991
QY 814 QMASVHQN-----AALD--SRWME--GFSNFOAFATKKEEY-----TNV---849
DB 992 KPAKVNNGVGVAGIDGTTRIDREIGFTGTNGSLDKSKPHLSKODINAGGKKITNIQSG 1051
QY 850 VIAXN-----VDFKFAEWGVTDFEMAPOY-----VSSTD 877
DB 1052 EIAQNSHDAVGGKLYDLKTELENKISSTAKTAQNSLHEFSVADQGNFTVSNPYSYD 1111
QY 878 GSFDSVIT---QNGYAFTRDYDLGISKPNKYGTADDLVKAICA-----LHSGKIKV- 925
DB 1112 TSKTSDVITFAGENG--ITTKVNGKQWVR-----VGIDQTKGLTTPKLTGVNNGKGIVID 1164

QY 926 -----MADWVPDMYAPPEKEVVVTRVDKYGTFPVAGSQIKNTLYVVDGKS 971
DB 1165 SONGNTITGLSNTLANVTNDK-----GSRVTEQGNIIKDEKTRAASIVDVL 1214
QY 972 SGKDOQ-----AKYGGAFLEELQAKYPFARKQISTGVPMDPDS---VKIK 1014
DB 1215 AGFNLQNGEAVDFVSTYDVTNFPADGNATTAKVTVDDTSKTSKVVDVNVDDTTIEVKDK 1274
QY 1015 QMSAKYFNGTNILGRGAG---YVLKQOAT-NYFNISD-NKEINFL-----PKTLLNQD 1063
DB 1275 KLGVK---TTTTSTGTGANKFALSNOQTGALVKASDIVAHLNLTSGDIQAKGASQAN 1331
QY 1064 SOVGF-SYDGKGYVYSTSG---YQAKNTFISEGDKWYFDDNNGYVMTGAOSING-----VN 1116
DB 1332 NSAGYVDADGNKVIYDSTDNKYQAKNDGTVDTKKEVAKDK---LVAQAQTPDGTGLAQN 1388
QY 1117 YFPLSNGIQLRDAILK---NEDGTVAYYVNDGRRYENGYYQFMSGVWR-----HFNNGEM 1168
DB 1389 VKSVINKEQVNDANKQGINEDNA-----FVKGLEKAAASDNKTNAAV 1431
QY 1169 SVG--LTVIDGOVYFDEMGYQAK--GKFVTTADGKIRYFDKQSGNMYNRPIENEEGKW 1224
DB 1432 TVGOLNVAQAQPLTAFAGTGTAKKLGSETLI-----KGGQTDITNKLTDNNIG-- 1479
QY 1225 LYLGEDGAATVGSQTINGQHLVFRANGVQVKGFEVTDHHRISYYDNGSGDQIRNRFVRN 1284
DB 1480 VVAGTDGTVTKLAKDLT-----NLNSVNAAGTKIDDK--GVSFVD--SSGQAKANTPVLS 1530
QY 1285 AQGWFYFDNNGYAVT-----GARTINGQLLYFRANGVQVKGFEVTDYGRISYYDNGSGD 1340
DB 1531 ANG-----LDLGGKVISNVGKGTQD-----AANVQQLNE-VRNLLGL-----GNAGN 1573
QY 1341 QIRNRFVNAQGWYF-----DNNGYA-----VTGARTI-----NGOHLVFRANGVQVK 1385
DB 1574 D-----NADGNQVNIADIKKDPNSGSSSNNRTVIKAGTVLGGKGNNDTEKLATGQIQVG 1626
QY 1386 GEFVTDHGRISYYDNGSGDQIRNRFVR-----NAQGWYFDNNGYAVTG 1431
DB 1627 ----VDK-----DGNANGDLSNVWVKTKQDKSKKALLATVNAAGQTYLNTN--PABA 1673
QY 1432 ARTINGQHL-YFRAN 1445
DB 1674 IDRNEQIRFEHVN 1688
RESULT 15
US-11-013-759-13
; Sequence 13, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; FILE REFERENCE: 1038-921MIS:jb
; CURRENT APPLICATION NUMBER: US/11/013.759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361.619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 13
; TYPE: PRT
; LENGTH: 1992
; ORGANISM: Moraxella catarrhalis
US-11-013-759-13

Query Match 2.8%; Score 219; DB 7; Length 1992;
Best Local Similarity 20.0%; Pred. No. 0.00013;
Matches 351; Conservative 215; Mismatches 665; Indels 524; Gaps 93;

QY 30 TISGLVKADS--NESKQISNDSTSVVTANESNVI-----TEATSK 71
Db 119 TLINDLINGHAVLKEIRSSKONDKYRRRTTASGHASTAVGAMSYAQHFHSNFAEGTRATAK 178
QY 72 QE-----AASQTNHVTITSSST-----SVNPKV 98
Db 179 SAYSLAVGLAATAEGOSTTAIGSDATSSSIALGALGATRAQLOQSIALOGGSVVTTQSDN 238
QY 99 VSNPYTVGTASNGEKLQNTTVDKTSSEAAANNISKQ--TTBADTDVDDSDNAANLOIL 156
Db 239 NSRPAYTPNTQALDPKFOATNTKAGPLSIGSNSIKRKIINVAGVKNKTDAVNVAQLEAV 298
QY 157 EKLPNVKEIDGKYVYDNG--KVTNFTLADGKILHDFEGATDTSIDTVNK-DIV 212
Db 299 VWAKERRIT---FOGDDNSTDKVIGLNTLTIKGGA---ETNALTDNNIGVYVKEADNS 351
QY 213 TTRSNIKKYNOVYDRSAOSFEHVDHYLTAESMYRPKYILKDGKTWTQSTEXKDFRLLMT 272
Db 352 GLVKLAKTLNLTENVTTLNATTVTKVGSSSTTABELLSLSLTFTQNT-----402
QY 273 WMPDQETQRY-----VNYMN-----AQLGINKTYDD-----TSNQL 304
Db 403 --GSQSTSVTVGVNGVKFTNNAETAAIGTTRI TRDKIGFARDGDVDEKQAPYLDKKQL 460
QY 305 QLNIAAATIQAKIEAKITLKNYDWLROTISAPVKTQSAWNS---DSEKPFDDHLONGAV 361
Db 461 KVGVAITTDNGIDAG-----NKKISNLAKGSSANDAVTIBQLKAAKPTLNAGAG 510
QY 362 LYDNEBKLPYANSYRIILNRT-PTNQTGCKOPRYTADNTIGGYEBELLANDVDSNPVVO 420
Db 511 ISVTPTEISVDAKSG--NVTAPYNI GVKTELNSDGTSDKFS---VKGSGTNNSLVT 563
QY 421 ABQL-NWLH-----FLMNFENIYANDPDANFDSIRVDADVNDVA-DLLOIAG-----465
Db 564 AEHLASYLNEVRTADALSALQSFVKEEDDDANAITVAKDTTKNAGAVSILKXKNGKLT 623
QY 466 -----DYLKAAKGHKND-----KAANDHLSILEAWSNDTPLYLHDDGN 505
Db 624 VATKKGDTVTFLGSQDSGLTIGKSTLNNDGLTVKOTNEQIQV-----GANGIKFTNVNGSN 679
QY 506 -----MINMDNKLRLSLFS-----LAKPLNQRSCM---NPLITNSLVNRTDDNAETAA 551
Db 680 PTGIANTARI TRDKIGFAGSGAVDTNPKPYLDQKLVGNVKIINTGIN-AGGKAITGL 738
QY 552 VPSYSFIRAHS---EVODLIADIK--AEINPNVVGYSFTMEIKKAFE-IYNKDL- 603
Db 739 SPTPLSIADQSSRNIELGNTIQDKOKSNAASIN-DILNTGFNLKNNNPIDFVSITVID 797
QY 604 -----ATEKKYTHYNALSYALLTNKSSVPRVYGDMPDQGYNAHKTIYNEAIELL 658
Db 798 FANGNATTATVTH-DTA-----NKTS--KVWY-DVNVD-----TTIHLTGTDNKK 839
QY 659 KARIKV-----SGQAMRNOQVGNSE-----IITSVRYGKGALKATD 696
Db 840 KLGVTITKLKNTSANGNTATNFVNSSDEDALVNAKOIAENLNTLAKEIHTTKG-----893
QY 697 TGDRTTRTSGAVIEGNPNRLKASDRVVVNMGAHQA VRPLLTTDNGIKAYHSDQ 756
Db 894 TADTALQTFVVKVDENN-----NADDANAITVGQKANNQVNTLTKEGELNI-----943
QY 757 EAAGLVRYNDRGELIF---TAADIKGYANPOVSGYLVGVVPGVGAALIKMFAIRLARPHQ 813
Db 944 -----KTDKNGTVTFGINTTSGLKAGKSTLNDGGLSIKNPTGSEIQVGA-----DGV 991
QY 814 QNASVHON---AALD--SRVWFE--GFSNFOAPATKKEEY-----TNV---849
Db 992 KPAKYNNNGVWAGIDGTTRITRDEIGFTGTNGSLDKSKPHLSKDCINAGGKKITNIQSG 1051
QY 850 VIANK-----VDKFAEWGVTDFEMAPOY-----VSSTD 877
Db 1052 EIAQNSHDAVTCGKIYDLKTELENKISSTAKTAQNSLHFSVADEQGNFTVSNPYSSVD 1111
QY 878 GSFLDSVI-----QNGYAFTDRYDLGLISKPNKYGTADDLVKAIKA-----LHSGIKV- 925

Db 1112 TSKTSDVITFAGENG--ITTKVNGKVVR-----VGIDQTKGLTTPKLTVGNNGKGIIVD 1164
QY 926 -----MADWVPDQMYAFPEKEVVVATRVKYGTPVAGSQIKNTLYVVVDGKS 971
Db 1165 SONGQWTTIGLSENTLANVTNDK-----GSVRTTEQGNIIKDDEKTRAASIVDVLS 1214
QY 972 SKGQOQ-----AKYGGAFLBELOAKYPELPFARKQISTGVPMDPS---VKIK 1014
Db 1215 AGFNLOQNGEAEVFSVYDTVNFADGNATTAKVTYDDTSKTSKVVDVNVVDDTTIEVKDK 1274
QY 1015 QMSAKYFNGTNIILGRGAG---YVLKDOAT-NYFNILSD-NKEINFL-----PKTLNQD 1063
Db 1275 KLGVK---TTTLTSTGTGANKFALSNQATGDALVKASDIVAHLNLTLSGDIQTAKGASOAN 1331
QY 1064 SQVGF-SYDGKGYVYVYSTG--YQAKNTFISGDKWYYPDNNGYMYVTGAQSLNG---VN 1116
Db 1332 NSAGYVDADGNKVIYDSTDNKYQAQNDGTVDKTEKADK---LVAQAQTPDGTTLAQMN 1388
QY 1117 YYFLSNGLQLRDAILK---NEDGTYAYYNGDGRRYENGYQFMSGVWR-----HFNNGEM 1168
Db 1389 VKSVINKEQVNDANKKQGINEDNA-----FVKGLEKAASDNKTKNAAV 1431
QY 1169 SVG--LTVIDGQVQVDEMGYQAK--GKEVTTADGKIRYPDKQSGSNMYNRNFIENEGKW 1224
Db 1432 TVGDLNVAQAQTPLTTPAGDTGTAKKLGELTI-----KGGQTDNTKLTDDNNIG-- 1479
QY 1225 LVLGSDGAAVTGSQTINGQHLRYFRANGVQVKGFEVTDHHRISYYDNGSDQIRNRFVN 1284
Db 1480 VVAGTDGFTVKLAKOLT-----NLNSVNAAGTKIDDK--GVSFVD--SSQQAQKANTPVLS 1530
QY 1285 AQGQWYFDNNGYAVT---GARTINGQLLYFRANGVQVKGFEVTDRYGRISYYDNGSGD 1340
Db 1531 ANG-----LDLGKVISNVGKGTKDTD-----AANVQQLNE-VRNLLGL-----GNAGN 1573
QY 1341 QIRNRFVNAQGWYF-----DNNGYA---VTGARTI-----NGQHLRYFRANGVQVK 1385
Db 1574 D-----NADGNQVNIADIKKDPNSGSSSNRTVIKAGTVLGGKGNNDTEKLTATGQV 1626
QY 1386 GEFVTDHGRISYYDNGSDQIRNRFVR-----NAQOWFYFDNNGYAVTVG 1431
Db 1627 ---VDK-----DGNANGDLSNVVVKTKQDKGSKALLATYNAAGQTYLNTN---PABA 1673
QY 1432 ARTINGOHL-YPRAN 1445
Db 1674 IDRINEQGI RPFVN 1688

Search completed: February 11, 2006, 20:57:47
Job time : 19.3936 secs

THIS PAGE BLANK (USPTO)

Result No.	Query	Score	Match	Length	DB	ID	Description
1	7741	100.0	1475	2	B3135	gtfB protein precu	
2	5285	68.3	1375	2	JT0345	dextranucrase (EC	
3	4531	58.5	1592	2	A45866	glucosyltransferas	
4	3839.5	49.6	1431	2	A45866	dextranucrase (EC	
5	3506	45.3	1577	2	T30858	glucosyltransferas	
6	3289	42.5	1599	2	S22737	glucosyltransferas	
7	3243	41.9	1508	2	T31098	probable dextranu	
8	3242	41.9	1290	2	JC5473	dextranucrase (EC	
9	3238.5	41.8	1449	2	T30857	glucosyltransferas	
10	3229.5	41.7	1449	2	T30552	glucosyltransferas	
11	3196.5	41.3	1518	2	A44811	glucosyltransferas	
12	2945.5	38.1	1365	2	A41483	glucosyltransferas	
13	553	7.1	2817	2	B97033	uncharacterized pr	
14	473	6.1	563	2	A37184	glucan-binding pro	
15	454	5.9	2710	2	A37052	toxin A - Clostrid	
16	398.5	5.1	648	2	S10869	enterotoxin A - Cl	
17	382	4.9	2367	2	S70172	toxin B - Clostrid	
18	381	4.9	2366	2	S10317	toxin B - Clostrid	
19	366	4.7	2178	2	S55805	alpha-toxin - Clo	
20	357	4.6	2564	2	I40884	cytotoxin L - Clo	
21	275	3.6	721	2	C97980	endo-beta-N-acetyl	
22	275	3.4	658	2	E95111	endo-beta-N-acetyl	
23	256.5	3.3	340	2	G95043	choline binding pr	
24	254.5	3.3	1335	2	T30211	autolysin E - Stap	
25	250.5	3.2	2893	2	A64556	toxin-like outer m	
26	250	3.2	1806	2	AF1717	probable peptidogl	
27	248	3.2	619	2	A97887	surface protein ps	
28	248	3.2	619	2	A41971	surface protein ps	
29	247	3.2	2902	2	C71953	toxin-like outer m	

```
Db 241 TAESVVRPKY ILKDGKWTQSTEKDFRPLLMTWMPDQETQRYVYVMNAQLGINKTYDDT 300
Qy 301 SNQLQNTAAATQAKIEAKITTLKNTWLRTOTISAFVKTSAMNSDSEKPPDDHLQNGA 360
Db 301 SNQLQNTAAATQAKIEAKITTLKNTWLRTOTISAFVKTSAMNSDSEKPPDDHLQNGA 360
Qy 361 VLYDNEGKLTYPVANSNRYLNRTPNTQCKKDPRTYADNTTGGYBFLANDVDNSNPVQ 420
Db 361 VLYDNEGKLTYPVANSNRYLNRTPNTQCKKDPRTYADNTTGGYBFLANDVDNSNPVQ 420
Qy 421 AEQLNLWFLMNFNIYANDPDANFDSIRVDADVNDADLLQIAGDYLKAAGIHKNDKA 480
Db 421 AEQLNLWFLMNFNIYANDPDANFDSIRVDADVNDADLLQIAGDYLKAAGIHKNDKA 480
Qy 481 ANDHLSILEAWSNDTPYLHDDGDMNINNDKRLSLFLSLAKPLNQSGMNPPLITNSLIV 540
Db 481 ANDHLSILEAWSNDTPYLHDDGDMNINNDKRLSLFLSLAKPLNQSGMNPPLITNSLIV 540
Qy 541 NRTDDNAETAAPVPSYFIRAHDSVQDLDIADIIKAEINPNVVGYSFTWEEIKKAFIYNK 600
Db 541 NRTDDNAETAAPVPSYFIRAHDSVQDLDIADIIKAEINPNVVGYSFTWEEIKKAFIYNK 600
Qy 601 DLLATEKKYTHYNTALSALLTNKSSPRVYVYGMFTDDGOYMAHKTINYEAIETLLKA 660
Db 601 DLLATEKKYTHYNTALSALLTNKSSPRVYVYGMFTDDGOYMAHKTINYEAIETLLKA 660
Qy 661 RIKYVSGGQAMNQVGNSEIITSVRYGKGALKATDGTGRTTRTSGVAVIEGNNPSLRK 720
Db 661 RIKYVSGGQAMNQVGNSEIITSVRYGKGALKATDGTGRTTRTSGVAVIEGNNPSLRK 720
Qy 721 ASDRVVVMGAHKNQVRPLLLTTDNGIKAVHSDQEAAGLVRYTNDRGELIFTAADIKG 780
Db 721 ASDRVVVMGAHKNQVRPLLLTTDNGIKAVHSDQEAAGLVRYTNDRGELIFTAADIKG 780
Qy 781 YANPQVSGYLGVPVPGVGAALIKWFALRLARPHQOMASVHQNAALDSRVMFEGFSNFQAPA 840
Db 781 YANPQVSGYLGVPVPGVGAALIKWFALRLARPHQOMASVHQNAALDSRVMFEGFSNFQAPA 840
Qy 841 TKKEBYTNVIAKNVDKFAEAGVTDPEMAPQVVSSTGDSFSDVITQNGYAFTRDRLDGLS 900
Db 841 TKKEBYTNVIAKNVDKFAEAGVTDPEMAPQVVSSTGDSFSDVITQNGYAFTRDRLDGLS 900
Qy 901 KPNKGTADDLVKAIKALHSKGIKVMADWPDMQYAFPEKEVVTATRVKDYKTPVAGSQI 960
Db 901 KPNKGTADDLVKAIKALHSKGIKVMADWPDMQYAFPEKEVVTATRVKDYKTPVAGSQI 960
Qy 961 KNTLVVDGKSGKDDQQAQYGGAFLEELQAKYPELFARKQISTGVPMPSVKIKOWSAKY 1020
Db 961 KNTLVVDGKSGKDDQQAQYGGAFLEELQAKYPELFARKQISTGVPMPSVKIKOWSAKY 1020
Qy 1021 FNGNTILGRGAGYVLKDAQNTYFNISDNKEINFLPKTLNLDQSDQVGSYDGKGYYVYST 1080
Db 1021 FNGNTILGRGAGYVLKDAQNTYFNISDNKEINFLPKTLNLDQSDQVGSYDGKGYYVYST 1080
Qy 1081 SCYQAKNTFISGDKWYFDNNGYVMTGAOSINGVNYFLSNGQLRLDAILKNEGDTYAY 1140
Db 1081 SCYQAKNTFISGDKWYFDNNGYVMTGAOSINGVNYFLSNGQLRLDAILKNEGDTYAY 1140
Qy 1141 YGNDGRRYENGYYQPMGSVWRHFNNGEMSVGLTVIDGQVQYFDENGYYQAKGKFTVTADGK 1200
Db 1141 YGNDGRRYENGYYQPMGSVWRHFNNGEMSVGLTVIDGQVQYFDENGYYQAKGKFTVTADGK 1200
Qy 1201 IRYFDKQSGMYRNRFIENEGKWLVLGEDGAATVTSQTINGQHLYFRANGVQVKGFEV 1260
Db 1201 IRYFDKQSGMYRNRFIENEGKWLVLGEDGAATVTSQTINGQHLYFRANGVQVKGFEV 1260
Qy 1261 DHGGRISYVDGNSGDOIERNRFRVNAQGWYFDNNGYAVTGARTINGQLLYFRANGVQVK 1320
Db 1261 DHGGRISYVDGNSGDOIERNRFRVNAQGWYFDNNGYAVTGARTINGQLLYFRANGVQVK 1320
Qy 1321 GEFVTDYGRISYYDGNSGDOIERNRFRVNAQGWYFDNNGYAVTGARTINGQHLYFRAN 1380
Db 1321 GEFVTDYGRISYYDGNSGDOIERNRFRVNAQGWYFDNNGYAVTGARTINGQHLYFRAN 1380
```

```
Qy 1381 GVQVKGFEVTDHGRISYYDGNSGDOIERNRFRVNAQGWYFDNNGYAVTGARTINGQHL 1440
Db 1381 GVQVKGFEVTDHGRISYYDGNSGDOIERNRFRVNAQGWYFDNNGYAVTGARTINGQHL 1440
Qy 1441 YFRANGVQVKGFEVTDYGRISYYDANSGERVRIN 1475
Db 1441 YFRANGVQVKGFEVTDYGRISYYDANSGERVRIN 1475

RESULT 2
JT0345
destranucrase (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)
N:Alternate names: sucrose 6-glucosyltransferase
C:Species: Streptococcus mutans
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 09-Jul-2004
C:Accession: JT0345; C3135
R:Ueda, S.; Shiroza, T.; Kuramitsu, H.K.
Gene 69, 101-109, 1988
A:Title: Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.
A:Reference number: JT0345; MUID:89137980; PMID:2976010
A:Accession: JT0345
A:Molecule type: DNA
A:Residues: 1-1375 <UED>
A:Experimental source: GS-5
A:Cross-references: UNIPROT:P13470; UNIPARC:UPI00000155515
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
J. Bacteriol. 169, 4263-4270, 1987
A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.
A:Reference number: A33135; MUID:87308013; PMID:3040685
A:Accession: C31335
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-349 <SHI>
A:Cross-references: UNIPARC:UPI000014E25D; GB:M17361
C:Genetics:
C:Gene: gtfC
C:Function:
A:Description: catalyzes the synthesis of both water-soluble and water-insoluble glucans
A:Keywords: duplication; glycosyltransferase; hexosyltransferase
F:1-34/Domain: signal sequence #status predicted <SIG>
F:35-1375/Product: glucosyltransferase #status predicted <MAT>
F:1126-1145/Domain: cpl repeat homology <CP1>
F:1253-1272/Domain: cpl repeat homology <CP2>
F:1318-1337/Domain: cpl repeat homology <CP3>

Query Match 68.3%; Score 5285; DB 2; Length 1375;
Best Local Similarity 75.9%; Pred. No. 3.4e-248;
Matches 1031; Conservative 101; Mismatches 169; Indels 58; Gaps 14;

Qy 1 MDKKVRYKLRKVKRWTVSVASAVMTLTTLSGGLVKADSNESKSIQINDSWTSVVTANE 60
Db 1 MEKKVRFKLRKVKRWTVSVASAVMTLTTLSGGLVKADSDTDDRQAVTESQASLVTTSE 60
Qy 61 -----ESNVITEATSKQEAASSQNTNHTVTSSSTSVVPKVVSNPTVCGETA-- 109
Db 61 AAKETLFTATDTSATSSQPTA--TVTDNVSTNQSTNTANTANFVVKPTTSEQAKT 118
Qy 110 SNBKLQNTQTTVDKTSSEA-----AANNISQTTTEAD-----TDVIDDSN 149
Db 119 DNSDKIITTSKAVNRLTATGKFPVANNNTAHPKVTDKIVIPKPKGLKQPSLSQDDI 178
Qy 150 AANIQLIEKLPNVKEIDGKYVYNNNGKVRNFTLIADGKILHFPDETGYTDTSDTVNK 209
Db 179 AA----LGNVKNIRKVGKYYKYEDGTLQKNYALNINGKTFFFDETGALESNTLPSKKG 234
Qy 210 DIVTT-RSNLYKKYQNVDRSAQSFHVHDVHLYTAESWVRPKYLLKDGKTKTQSTSEKFRP 268
Db 235 NITNDNTNFAQNNQVYSTVDVANFHVHDVHLYTAESWVRPKYLLKDGKTKTQSTSEKFRP 294
Qy 269 LLMTWVDPQETQRYVYVMNAQLGINKTYDDTTSNQLNLNIAAATQAKIEAKITLKNTD 328
Db 295 LLMTWVDPQETQRYVYVMNAQLGINKTYDDTTSNQLNLNIAAATQAKIEAKITLKNTD 354
```

QY	329	WLQRTISAFVKTQSAWNSDSSEKPPDDHLQNGAVLYDNEGKLTYPYANSYRILNRTPTNQ	388
DB	355	WLQRTISAFVKTQSAWNSDSSEKPPDDHLQKGLLYSNNSKLTQSAWSYRILNRTPTNQ	414
QY	389	GKKDPRYTADNTIGGYEFLANDVNSNPVQAEQLNLHLFIMFNGIYANDPDANPDSI	448
DB	415	GKKDPRYTADRTIGGYEFLANDVNSNPVQAEQLNLHLFIMFNGIYANDPDANPDSI	474
QY	449	RVDVNDVNDADLLQIAGDYLLKAAKGIIKXDKAANDHLSILEAWSNDPTPYLHDDGDNNIN	508
DB	475	RVDVNDVNDADLLQIAGDYLLKAAKGIIKXDKAANDHLSILEAWSYNDPTPYLHDDGDNNIN	534
QY	509	MDNKLRLSLLPSLAKPLNQRSGMNPPLIITNSLVNRTDDNAETAAPVPSYSFIRAHUSEVODL	568
DB	535	MDNRLURLSLLSLAKPLNQRSGMNPPLIITNSLVNRTDDNAETAAPVPSYSFIRAHUSEVODL	594
QY	569	IADIIKASINPNVVGYSFTTBEIEIKAFPIYKNDLLATEKKYTHYNTALSYALLLTNKS	628
DB	595	IRNIIRTEINPNVVGYSFTTBEIEIKAFPIYKNDLLATEKKYTHYNTALSYALLLTNKS	654
QY	629	PRVYVYGMFTDDGGYMAHKTINYEAIEPLLKARIKYVSGQAMRNQVGNSEIITSRYG	688
DB	655	PRVYVYGMFTDDGGYMAHKTINYEAIEPLLKARIKYVSGQAMRNQVGNSEIITSRYG	714
QY	689	KGALKATDGTGRTTRTSGAVIEGNNPSRLRKASDRVVVNNGAAHQNQAYRPLLLTTDNG	748
DB	715	KGALKATDGTGRTTRTSGAVIEGNNPSRLRKASDRVVVNNGAAHQNQAYRPLLLTTDNG	774
QY	749	IKAVHSDOEAAGLVRYTNDRGELIFTADI IKGYANPQVSYGLVGVVPGVGAALIKWFA	808
DB	775	IKAVHSDOEAAGLVRYTNDRGELIFTADI IKGYANPQVSYGLVGVVPGVGAALIKWFA	832
QY	809	A---RPHQOMASVHQNAALDSRVMFEGFSNFQAFATKKEEYTNVVIKQVDFKFAEWG	865
DB	833	AASTAPSTDGKSVHQNAALDSRVMFEGFSNFQAFATKKEEYTNVVIKQVDFKFAEWG	892
QY	866	FEMAPQVYSSPDGFSFLDSVIONGYAFTDRYDLGISKPNKYGTADDLVKAIKALHSKGI	925
DB	893	FEMAPQVYSSPDGFSFLDSVIONGYAFTDRYDLGISKPNKYGTADDLVKAIKALHSKGI	952
QY	926	MADVDPQMYAFPEKEVVTATRVDKYGTGPVAGSQIKNTLYVVDGKSGKQQAQKYG	985
DB	953	MADVDPQMYALPEKEVVTATRVDKYGTGPVAGSQIKNTLYVVDGKSGKQQAQKYG	1012
QY	986	EELQAKYPELPARKQISTGVPMDPVSKIKOWSAKYFNGTILGRGAGVYLKDAQATNTY	1045
DB	1013	EELQAKYPELPARKQISTGVPMDPVSKIKOWSAKYFNGTILGRGAGVYLKDAQATNTY	1072
QY	1046	-ISDNKEINFLPKTLN-----QDSQVGSYDGRGYVYVYSGYQAKNTFISEGDKWY	1099
DB	1073	LVSDN---TFLPKSLVNPNHGTSSTVGLVFDGKGYVYVYSGYQAKNTFISEGDKWY	1129
QY	1100	DNNGYMTGAOSINGVNYFTLSNGLQRLADALLKNEEDGTAYYNGDGRYENGYYQFM	1159
DB	1130	DNNGYMTGAOSINGVNYFTLSNGLQRLADALLKNEEDGTAYYNGDGRYENGYYQFM	1188
QY	1160	WRHFNNGEMSVGLTVIDQVOYEPDMGYQAKGKFTTADGKIRYFEDKSGNMWYENR	1219
DB	1189	WRYFQNGHMAVGLTRVHGAVOYFASGFGQAKGQFITTADGKLRFYFDRSDSGNQIS	1248
QY	1220	EEGKLYLGEDCAAVTSQTINGOHLVFRANGVOVKGEFVTDHHRISYYDNGSGDQIRN	1279
DB	1249	SKGEWFLFDHNGVAVTGVTFNGQRLVFKPNGVQAKGEFTRDANGLYRLYDYPNSG	1308
QY	1280	RFVNAQOWFYFDNNGYAVTGARTING-----QLLYFR 1313	
DB	1309	RFVNSKGWFLFDHNGIATVGARVNVGHASILSLWVFR 1347	
RESULT 3			
A38175			
glucosyltransferase precursor - Streptococcus sobrinus			

C.Species: Streptococcus sobrinus
C.Date: 28-Aug-1992 #sequence_revision 28-Aug-1992 #text_change 15-Mar-2004
C.Accession: A38175
J.Abdo, H.; Mtsamura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.
J. Bacteriol. 173, 989-996, 1991
A>Title: Peptide sequences for sucrose splitting and glucan binding within Streptococcus
A.Reference number: A38175; MUID:91123227; PMID:1704006
A.Accession: A38175
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-1592 <ABO>
A.Cross-references: UNIPARC:UPI000012BCB2; GB:D90213; NID:g217032; PIDN:BAAL4241.1; PID:
F.1033-1112/Domain: cpl repeat homology <CP1>
F.1232-1241/Domain: cpl repeat homology <CP2>
F.1287-1306/Domain: cpl repeat homology <CP3>
F.1330-1351/Domain: cpl repeat homology <CP4>
F.1352-1371/Domain: cpl repeat homology <CP5>
F.1402-1420/Domain: cpl repeat homology <CP6>
F.1465-1484/Domain: cpl repeat homology <CP7>
F.1513-1532/Domain: cpl repeat homology <CP8>

```
QY 706 GVAVTEGNPSRLKASDRVVNNMGAHKNQAYRPLLLTTDNGIKAYHSDQEA--AGLVR 763
Db 702 GGVVMGNQPNPSLDGK-VVALNMGAHANQBYRALMVSTKDGVAITYATDADASKAGLVK 760
QY 764 YTNDRGELIFTAADIKGVANPOVSQYLGWVVPVGA--LIKMFALRLARPHQWASVHQ 820
Db 761 RTDENGILYFLNDLLKGVANPOVSQYLGWVVPVGAADQDQIRVASDSTADTGK--SLHQ 818
QY 821 NAALDSRVMEFSGFNQFATKKEBYTNVIAQVNDKFAEAGVTDPEMAPQVVSSTDSGF 880
Db 819 DAAMDSRVMEFSGFNQFATKKEBYTNVIAQVNDKFAEAGVTDPEMAPQVVSSTDSGF 878
QY 881 LDSVLQNGVAFTRDYLGISPKNYKTADDLKAIKALHSKGIKVMADVPDQWAFPEK 940
Db 879 QDSVIQNGVAFTRDYLGISPKNYKTADDLKAIKALHSKGIKVMADVPDQWAFPEK 938
QY 941 EVVTRATRVKGYTPVAGSQIKNTLVVVDGKSGKQQAQYGGAFLEELQAKYPELFAFKQ 1000
Db 939 EVVTRTRDKFGKPIAGSQINHSLVYDTKSGDDYQAKYGGAFLEELQAKYPELFAFKQ 998
QY 1001 ISTGVMPDPSVKIKQWAKYFNGTILGRGAGYVLKQATNTYFNISDNKEINFLPKTL 1060
Db 999 MSTQAIQDPSVKIKQWAKYFNGTILGRGAGYVLKQATNTYFNISDNKEINFLPKTL 1056
QY 1061 NQDSQVGSFYDQKGVVYVYST-SGYQAKNTPTISEGDKWYFDDNNGYMTGAQSGINGVYF 1119
Db 1057 GKWESGTRYDQKGVVYVYSSATGDQVQKASFTIAGNLVYFGDKGYMTGAQSGINGVYF 1116
QY 1120 LNSGLQRLDALIKNBDGTVAAYGNDGRRY--ENGYQFMGSGVWRHFNNGEMSVGLTVIDG 1177
Db 1117 LENGTALRNTIYTDAGNSHYANDGKYENENGVOQP-GNDWRYPKQGNMAVGLTVDG 1175
QY 1178 QVQYFDEMGYQAKGFVTTADGKIRYFDKQSGNMYRNRPFIENEGKWLVLGEGDAAVTGS 1237
Db 1176 NVQYFQKGVQAKDKIIITRDGKRYFQDNGNAVTNTFIADKTGHVYVLGKGVAVTGA 1235
QY 1238 QTINGOHLYFRANGVOVKGFEVTDHGRISYYDNGSGDQIRNRFVRNAQGWYFDDNGY 1297
Db 1236 QTVGQKLYFEANGQVQKGFVTSDEGLYFYDVGSDGMDWTDFTIEDKAGNMYFLGKQGA 1295
QY 1298 AVTGARTINGQLLYFRANGVOVKGFEVTDHGRISYYDNGSGDQIRNRFVRNAQGWYF 1357
Db 1296 AVTGARTINGQLLYFRANGVOVKGFEVTDHGRISYYDNGSGDQIRNRFVRNAQGWYF 1355
QY 1358 DNGYAV-----TGAR 1368
Db 1356 GNDGVADPSPVVKGTQFKDASGALRYNLLKGQLVTGSGWYETANHDWVYIQSGKALTGEQ 1415
QY 1369 TINGOHLYFRANGVOVKGFEVTDHGRISYYDNGSGDQIRNRFV----- 1412
Db 1416 TINGOHLYFKDGHQVKGQLVTGTDGKRYYYDANGSDQAFNKSVTYNGKTYFFGNDGTAQ 1475
QY 1413 -----RNAQGWYFDDNGYAVTGARTINGOHL 1440
Db 1476 TAGNPKGTQFKDGSIRFYSMEGQLVTGSGWYVNAQGWLYV-KNGKVLTLGLQTVGSORV 1534
QY 1441 YFRANGVOVKGFEVTDHGRISYYDANGSERV 1472
Db 1535 YFDENGIQAKGKAVRTSDGKIRYFDENSGSMI 1566

RESULT 4
A:Accession: A45866
A:Title: precursor - Streptococcus mutans
C:Species: Streptococcus mutans
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Mar-2004
C:Accession: A45866
R:Honda, O.; Kato, C.; Kuramitsu, H.K.
J. Gen. Microbiol. 136, 2099-2105, 1990
A:Title: Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the glucosyl
A:Reference number: A45866; MUID:91100958; PMID:2148600
A:Accession: A45866
A>Status: preliminary
```

```
A:Molecule type: DNA
A:Residues: 1-1431 <HON>
A:Cross-references: UNIPARC:UPI000017AC5C; GB:M29296
C:Keywords: glycosyltransferase; hexosyltransferase
F:181-201/Domain: cpl repeat homology <CP1>
F:1127-1146/Domain: cpl repeat homology <CP2>
F:1192-1211/Domain: cpl repeat homology <CP3>
F:1257-1276/Domain: cpl repeat homology <CP4>
F:1277-1297/Domain: cpl repeat homology <CP5>
F:1321-1340/Domain: cpl repeat homology <CP8>
F:1341-1361/Domain: cpl repeat homology <CP6>
F:1385-1404/Domain: cpl repeat homology <CP7>

Query Match 49.6%; Score 3839.5; DB 2; Length 1431;
Best Local Similarity 53.1%; Pred. No. 4.2e-178;
Matches 765; Conservative 231; Mismatches 393; Indels 53; Gaps 22;

QY 1 MDKRVYKLRVKRWTVSVASAVMTL-TTTLGGGLVKADSNESK-----QISNDSNTS 54
Db 1 METKRYKMKVKRWTVSVASGLITLGTTLTGLSSVSAETEQOTSQTSKVVQKSEDDKAA 60
QY 55 VVTANEESNVITEATSKQEAASSQTNHTVTSSSTSVVNPKEVSVNPYTVGETASNGEK 114
Db 61 SESSQTDAPKTKQAQTEQTAQSQAN-VADTSTSTIKETPSQNIITQANSDDKKTVTNTKS 119
QY 115 LQNTQTV-DKTSAAANNISKOT-TEADTVDIDDSNAA-----NLQILEKLPNKYKIDG 167
Db 120 EEAQTSERTKQSEEAQTASSQALTOAKBELTKQRTAAQENKPNVDLAAIPNKYQIDG 179
QY 168 KYVYDNGKVRNFTLIADGKIHLHFD-TGAYDTSDTIDVTNKQIVTTRSNLYKKYNQVY 226
Db 180 KYVYIGDQPKKFNALTVNNKLVYFDKNTGALTDTDS-QYQFKQGLTKLNNDYTHNQIV 238
QY 227 DRSAQSPHVDHYLTAEWSYRPKVILKDGKWTOSTEKDFRPLMTWMPDQETQRYVNY 286
Db 239 NFENTSLETIDNYVTADSWYRPKDILKNGKWTWASSSEDLRPLMTWMPDKQTQIAYLNY 298
QY 287 MNAQ-LGINKTYDDTSNQLNIAAATQAKIEAKITTLKNTDMLRQIISAFVKTSQAWN 345
Db 299 MNQOGLGTGENYTDSSQESLNLAQTQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 358
QY 346 SDSEKPPD---DHLQNGAVLYDNEGKLTVPYANSNYRILNRTPTNQTCKDPRYTADNTI 401
Db 359 SQTSDTSAGKDHLOGGALLYNSDK-TAYANSDYRLNLTPTTSQTK--PKYFEDNSS 415
QY 402 GGYEFLANDVDSNPVQAEQNLNHLFNLNMGNIYANDPDANFDSIRVDAVDNVDADLL 461
Db 416 GGYDFLLANDIDNSNPVQAEQNLNHLNMGYSIVANDPEANFQGVVDAVDNVDADLL 475
QY 462 QIAGDYLKAAKGIHKNKDKAANDHLSILEAGSDNDTPYLHDDGDNMNMNKLRLSLPSL 521
Db 476 QIASDYLKAHYGVDPKSEKNAIHLNLSILEAGSDNDPQYNKDTKGAQLPINDKLRLSLYAL 535
QY 522 AKPLNQ-----RSGMNPILTNSLVNRTDDNAETAAPVSPYFIRAHDSVQODLIADI 573
Db 536 TRPLEKASKNKNEIRSGLEPVITNSLNRSAGKNSERMANVYIFIRAHDSVQTVIAKII 595
QY 574 KAEINPNVYVGSFTMEIEIKKAFIYNKOLLATEKKYTHYNTALSAYALLTNKSSVPRVY 633
Db 596 KAOINPKTDGLTFTLDELKQAFKIVNEDMRQAKKYTQSNIPYATAVALMLSNKSIDTRLY 655
QY 634 GDMFTDDQYMAHKTINTYEALITLKKARIKVVSGQAMRNOOVGNSE-----IITS 684
Db 656 GDMYSDDDQYMATKSPYYDAIDTLKKARIKAAAGGQDMKITTYVEGDGSHMDWDYTGVLTS 715
QY 685 VRYGKALKATDTGDRTRTSGAVIEGNNPSRLKASDRVVNNMGAHKNQAYRPLLT 744
Db 716 VRYGTANEATDQSGEATKQGMVITSNNSLKNQNDKVIIVNMGAAHKNQAYRPLLT 775
QY 745 TDNGIKAYHSDQEAAGLVRYTRNDRGELIFTAADIKGYANPOVSQYLGWVVPVGAALIKMF 804
Db 776 TKDGLTSYTSDAAKSLYRKNTDKGELVFDASDIQGYLNPQVSGYLAVMVVPVGS--DNQ 833
```


Db 1219 YFYDNGYQAGVFPVPTKDGHLMPFGCDGSGERYKSGFFE-QDGNWYVANDKGYVATGFTKV 1277
QY 1236 -----GSGTINGQHLVFRANGVQVKGFEV 1259
Db 1278 GKQNLVFNKGVQVKNRPFQVGDATYANNNEGDLVLRGAQTINGDELYFDESQGVKGFEV 1337
QY 1260 TDHGHGRISYDNGSGDQIRNR-----FVRNAQGWQFYFDNN 1295
Db 1338 NNPDPGTTSYDAITGVKLVDTSLVVDGQTFNVDAKGAVVTKAHTPGFYTTGDDNNWFYADSV 1397
QY 1296 GYAVTCAPTINGQGLVFRANGVQVKGFEVTDYGRISYDNGSGDQIRNRFVRNAQGWQF 1355
Db 1398 GRNVTGAQVINGQHLVFDANGQVKGFEVNTDGSRSFYHWTGDKLVSTFFATGHDWY 1457
QY 1356 YFDNNGYAVTGARTINGQHLVFRANGVQVKGFEVTDHGRISYDNGSGDQIRNRFVRNA 1415
Db 1458 YADDRGVVTVGAQVINGQKLFDDTGKQVKGAFATNANGSRSYHHWNTGNKLVSTFFTS 1517
QY 1416 QGWQFYFDNNGYAVTGARTINGQHLVFRANGVQVKGFEVTDYGRISYDNGSGDQIRN 1471
Db 1518 DNNWYADAKGVVVGEQINGQHLVFDQTKQVKGATATNPDGSISYDVHTGEK 1573

RESULT 7
T31098
probable dextranucrase (EC 2.4.1.5), extracellular - Leuconostoc mesenteroides
C:Species: Leuconostoc mesenteroides
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T31098
R:Monchois, V.; Renaud-Simeon, M.; Monsan, P.; Willemot, R.M.
FEMS Microbiol. Lett. 159, 307-315, 1998
A:Title: Cloning and sequencing of a gene coding for an extracellular dextranucrase (DS
A:Reference number: Z20981; MUID:98164374; PMID:9503626
A:Accession: T31098
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1508 fMON>
A:Cross-references: UNIPROT:052224; UNIPARC:UPI00000BB69B; EMBL:AF030129; NID:G2766611;
A:Experimental source: strain NRRL B-1299
C:Genetics:
A:Gene: dsrB
C:Function:
A:Description: produces dextran composed only of alpha(1-6) glucosidic bonds
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 41.9%; Score 3243; DB 2; Length 1508;
Best Local Similarity 43.7%; Pred. No. 3.7e-149;
Matches 682; Conservative 241; Mismatches 452; Indels 184; Gaps 32;
QY 2 DKKVRKLRKVRKRWTVS-VASAVMTLTTLGGGLVKADSNESKSQ-----ISNDNSTSV 56
Db 6 ERNRKLYKSKSVIGGLIUSTLMSWTATSQNVNADSTNTVTDKSVTVSSNNTT-- 63
QY 57 TANEESNVITEATSKQEAASSQTNHTVTTSSTSVNPKYVSNPYTVGETASNGEK-- 114
Db 64 --NQHDTVVDKQTI--PVKNDQTTQOIAANAQAQKVKASDTTDTQKAETANNTKED 119
QY 115 -LQNOT-----TTVDKTSAAANN----- 132
Db 120 SIDNLTKQLPAVPTANOKTGYLEKDGKWWYVTSNTLAKGLTTVDNHHQYFDNNGVQAK 179
QY 133 -----ISKOT-----TEADTDVDDSN 149
Db 180 GQFVTDNKTYYLDPNNGNAVTVGIIQOIGSQTALFNDNGEQVPADPYTPADGKTYFYDDKG 239
QY 150 AANLQILEKLPNVKEIDGKYIYDNGKVRNTFTLIADKILHFDETGAYTDTSDTVNK 209
Db 240 QATI-----GLKAINGHYFDSGLQKKGFTGVIDQVRYFDQESQGVSTTDSQIK 292
QY 210 DIVTTRSNLYKKNVYDRSAQSFEHVDHLYTAESWYRPKYLLDKGKTWTOSTEKDFRL 269
Db 293 EGLTSQTTDYTAHNAVHSDADFNFNGLTASSWYRPKDVLRNGQHWAEATTANDRP 352

QY 270 LMTWMPDQETQOYVNYMNAQLGI---NKTYDDTSNQLNLIAAAATQAKIBAKITTLKN 326
Db 353 VSWWFPKQTOVNYLNYM-SQGLIDNRQMFSLKQNAMLNACTTVQQAIEKIGVANS 411
QY 327 TDWLQRTISAFVKTSAMNSDSEKPPDDHLQNGAVLYDNEGKLTPTYANSNYRILNRPPTN 386
Db 412 TAWLKTAIIDDFIRTQPMNMSSEDPKNDHLQNGALTFFVN-SPLTPDTNSNFRLLNRPPTN 470
QY 387 QTKKDPRTADNTTGGYEFILLANDVNSNPVQAEQLNHLHFLMNFHNIYANDPDPANED 446
Db 471 QTVG--PKYTIIDQSGKFELLANDVNSNPVQSEQLNHLHFLMNFHNIYANDPDPANED 528
QY 447 SRVDAVNVDADLLQIAGDYLKAAGIKHNDKAAANDHLSILEAWGNDPTPYLHDDGDNN 506
Db 529 GIRVDAVNVDADLLQIADYFKAAYGVNDKNDATANQHLSDILEWHDNDPEYVKDFGNQ 588
QY 507 INMDNKLRLSLFSLAKPLNQRSGMNPITNSLVNRTDDNAETAAPVSPYSFIRAHDSEVQ 566
Db 589 LTMDDDYMHQTQLIWSLTCKMRMGTTMORFMDYYLVNVRNHDSTENTAIPNYSFVRAHDSEVQ 648
QY 567 DLIAIDIIKAEINPVV-GYSFTMEBIKKAPEIYNKDLATEKIKYTHYNTALSVALLLTNK 625
Db 649 TVIAQII-SELHPDVKNSLAPTADQLAEAFKIYNNDKQADKKYTOYNNPMSAYAMLLTNK 707
QY 626 SSVPRVYVYGDMDFTDDQYMAHKTINVEATETLLKARIKYVSGQAMRNQOVGSEIITSV 685
Db 708 DTVPRVYVYGDLYDDQYMAKSPYFPAIDNGLLKSRIKYVAGQSM---AVQNDILTNV 764
QY 686 RYKGALKATDGTDRTRTSGVAVIEGNNPNSURLKASDRVVVNMGAHQAQYRPLLLTT 745
Db 765 RYKGAMSVTSGNADTRTQIGVIVSNKENLALKSGDVTVLHMGAAHQAQYRPLLLTT 824
QY 746 DNGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKGYANPOVSGYLGWVVPVGAALIKMFA 805
Db 825 ADNLSYDNDNAP---VKYTNDDGLIDFNTBIYGVNRPQVSGFLAVVWVPGAD-----S 876
QY 806 LRLAR-----PHQOMASVHQAALDSRVMPGFSNFQAFATKCEYTNVIAKNVDKFA 859
Db 877 HQDARTLSDTAHHDGKTFHNSNALDSQVIEGFSNFQAFATNEDYTNVIAKNQQLPK 936
QY 860 EMGVTDPEMAPQYVSSDTGSLDSVIQNGYAFTRDYDLGISKPNKYGTADDLVKAIALH 919
Db 937 DWGITSFQLAPQYRSSTDTSLDSIIQNGYAFTRDYDLGYGPTKYGTVDQLRDAIKALH 996
QY 920 SKGIKVMADVPDQMYAPPEKEVVTATRVKYGTPVAGSQIKNTLYVVDKSGSGKQQAQ 979
Db 997 ANGIOAIADWVPDQIYNLPQGLATVTRTNSYGDKTNSDIDQSLYVQSRGGK-YQAQ 1055
QY 980 YGGAFLEBLOAKYPELPARKOISTGVPMDDPSVKIKQWSAKYFNGTNILGRGAGYVLKQQA 1039
Db 1056 YGGAFLEBLOAKYPELPARKOISTGVPMDDPSVKIKQWSAKYFNGTNILGRGAGYVLKQQA 1115
QY 1040 TNYTFNI-SDNKEINFLPKTLINODSQVGSFDGKGYVYVYSTGYQAKNFTISEG-KWY 1097
Db 1116 TDQYVYVTSNNNRDDELQKLTDDLSETFVDRDNGWVYVYLSGLYARNTFQDNGNY 1175
QY 1098 YFDNNGYVWVTAQSGINGVNYVFLSNGLOLRDALIKNEDGTAYYNGDGRYRNGYVQFMS 1157
Db 1176 YFDSTGHLVTFQNIINHHYFPLNGIIBLVOSFLQVADGSTTYFDQKGRQVENQYITDQT 1235
QY 1158 GWRHF-NNGEM-SVGLTVIDQVOYFDBMGVQAKGFVTTADTKIRYFDKSGMYNR 1215
Db 1236 GTAYFYQNDGTWTSFGFTIDGHKQFYKNGTQVKGQFVSDTDDHVFYLEAGNVA 1295
QY 1216 FIENEGKWLXIGEDAAVTSQTINGQHLVFRANGVQVKGFEV-DHH----- 1263
Db 1296 FAQNSQGWFLYDNDGIALTGLTINGVQNYFYADGHQSKGDFITIQNHVLTNPITGAI 1355
QY 1264 -----GRISYDNGSGDQIRNRFVRNAQGWQFYFDNNGYAVTGARTINGQHLVFRAN 1315
Db 1356 TTGMOQIGDKIFVFD-NTGNMLTNOYOTLDQWHLSTQGPADTGLVNLINGNLKYFQAN 1414

Query Match		41.8%;	Score 3238.5;	DB 2;	Length 1449;
Best Local Similarity		46.0%;	Pred. No. 5.7e-149;		
Matches 702;		Conservative 215;	Mismatches 429;	Indels 179;	Gaps 37;
QY	1	MDKKVRYKLRVKRWVTVSV-----ASAVMTLTLSGLVKADSNESKS-----	45		
Db	1	MDKKVHYKWKVKQWVTVIAVGLSLGAVSAVSLGT-NDGVVQADSDHTDATVAIPDITVD	59		
QY	46	--QISNDSNTSVVTANEESNVITEATSKQEAASSQTNHT-----VTTSSSSTSVNPKVYV	99		
Db	60	TGTVSNDT-----TAAQDPTTAAATN--DVATDQATPTATPDLTDTTNTVAANAVIDV	112		
QY	100	SNPYTVG-----ETASNGEKLQ--NQTITVDKTS--EAA-----	129		
Db	113	A---TVGTDRAATNTDNTATNDTAVDTTNNNTTDTTVDTRAATERRATGARRPTGG	169		
QY	130	--ANNISKOTTEAD--TDVIDDSNAANLQILEKLPNVKEIDGKYYYIDNNGKVRNFTL	184		
Db	170	RRATPVNGTNNANNTVTVVNDLDPATNNVVDGSPSHIKTNGKQYVVEDDGTIRKNYYL	229		
QY	185	-----IADGKILHFDGTAYTDTSDITVNKDIIVTTRSNLYKKYNQVYDRS	229		
Db	230	ERIGGSOYFNAETGELSNOKEYPDKNGG-TGSSADSTNTNTVNGDK-----NAPYGT	283		
QY	230	AQSFPHVDHYLTAESWYRPKYILKDGKTWTQSTEDKFRPLMTWMPDQBTQRYVNYMA	289		
Db	284	DKDIELVDGYFTANTWYRPKEILKDGKEWTASTENDKRPLLTVMWPSKAIQASLYNMYKE	343		
QY	290	Q-LGINKTYDDT'SNOQLNIAAATQAKIEAKITLTKNTDMLROTISAFVKTQSAWNSDS	348		
Db	344	QGLTNGQTYTSSQTQMDQNALEVQKRIEERIAREGNTDMLRTIKNFVKTPQPGWNSTS	403		
QY	349	EK-PFDDHLQNGAVLYDNEGKLTVPYANSNYRILNRTPTNOTGKDPRTYADNTIGYEL	407		
Db	404	ENLDNDNHLQGGALLYNDSR-TSHANSDYRLNLTPTTSQTKGNPKYTKDTSNGGPEPL	462		
QY	408	LANDVDNSPVVQABOLNWLHFLMFGNIYANDPDANFDSIRVDVNDVADLLQIAGDY	467		
Db	463	LANDIDNSNPVQABQLNHLWYIMNIGTITGGSDEDFGVRVDAVDVNVNADLLQIADY	522		
QY	468	LKAAAGIHKDKAANDHLSILEAKSDNDTPYLHDDGDNMNMNOKLRLSLFLSLAKPLNQ	527		
Db	523	FKAKYGADQSQDAQATKHSILEAWSHNDAYNEDTKGAQLPMDDPMLHALVYLSLRPTGN	582		
QY	528	RSQMPLTSLNVRTDDNAETAAPVPSYFIRAHDSVQDLADIKAENINNVVGYST	587		
Db	593	RSGEPLSLNSLNDRESGSKRWANVAFVRAHDSVQSIQIQLKNEINPQSTGNTFT	642		
QY	588	MEEIKKAFIYNKDILLATEKKYTHYNTALSYALLTNKSSVPRVYVYGMFTDDGQYMAHK	647		
Db	643	LDEMKAFEIYNKDMRSANKQYQYNIPSAYALMLTHKDTVPRVYVYGMFTDDGQYMAQK	702		
QY	648	TINYEAETLLKARIKYVSGGQAMRNQOVNSE-----LIITSRYGKGALKATDTGD	699		
Db	703	SPYDAIETLLKGRIRYAGGQDMKVNIGYNTNGWDAAGVLTSVRYGTGANSASDTGT	762		
QY	700	RTTRTSGVAVIBGNPNPSLRKASDRVVMGAAHKNQAVRPLLLTTDNGIKAYHSDQAA	759		
Db	763	AETRNOGHAVIVSNQOPALRL--TSLNTINMGAAHNRQAVRPLLLTNDGVATYLNDSAN	820		
QY	760	GLVRYTNDRGELIFTAADIKGYANPQVSGYGLGVWVPVGAALIKFALRLARPHQOMAS--	817		
Db	821	GIVKYTDGNGNLTPSANEIRGNRPQVDGLAVWVPVGAS--ENQDVRVAPSKERNSSGL	878		
QY	818	--VHQNAALDSRVWFGFNGFQAFATKBEYTNVJIAKVDKFAEAGVTDTEWAPQYVST	876		
Db	879	VYESNAALDSQVYIEGFSNFQDFVQNPQYTNKKIAENANLFPKSGITSEFEPQYVSSD	938		
QY	877	DGSLFDSVIQNGYATPDRYDLGISKPNKYGTADDLVKALKHLSKIKVMADWPDQMYA	936		
Db	939	DGSLFDSVIQNGYATPDRYDLGMSKDNKYGLSLADLKAALKHLSHAGVISAIADWPDQIYN	998		
QY	937	FPEKEVWATRVDKYTPVAGSQIKNTLYVVDGKSSGKDDQQAQKYGGAFLLELQAKYPELF	996		

Db	999	LPQDEVVTVATRVNNYGETKDGAIIDHSLYAAKTRTFGNDYQKYGGAFLDELKRLYPQIF	1058	
QY	997	ARQIQISTGVPMDPSPVKIKOWSAKYFNGTNIILGRGAGYVLKQDQATNTVFNISDNKEINF	1056	
Db	1059	DRVQISTGKRMVTTDEKITQWSAKYMWGTNILDGRSEYVLKN-GLNGYYGTNGKV--SLP	1115	
QY	1057	KTL-----LNQDSQVGFSDYDGKGV-----YYSTSGYQAKNTFISEGD-KWYFFDNN	1102	
Db	1116	KVVGSNQSTNGDNQNG---DGSCKFEKRLFSVRYRYNNGQYAKNAFIKNDGNGVYFFDMS	1172	
QY	1103	GYVMTGAOSINGVNYFPLSNGLQLRDAILKNEBDGTAYVYGNDRGRYENGYYQPMGVMWRH	1162	
Db	1173	GRMAVGEKTI DGKQYFFLANGVQLRDGYRON-----RR-----	1205	
QY	1163	FNNGEMSVGLTVIDSQVQYFDEM--YQAKGKFTVTADGKIRYFDKOSGNMYRNRFIENEE	1221	
Db	1206	-----GQVFYDQNGVLNANGK-----QDPKPDNNNNSG---RNQFVQIGN	1244	
QY	1222	GKWLVLGDBGAAVTSQTINGQHLYFRANGVQVKGEFVTDHHRISYYDGNSGDQIRNRF	1281	
Db	1245	NWVAYYDNGKRVTHQNINGQELFFDNGVQVKGRVTNE-NGAIRYYDANGSEMARNRF	1303	
QY	1282	VRNAQGFYFDNNGYAVTGARTINGQLLYFRANGVQVKGEFVTDYGRISYYDGNSGDQ	1341	
Db	1304	AEIEPGVWAYFNNDGTAVKGSQINGQDLYFDQNGRQVKA-LANVDGNLRYYDVNSGEL	1362	
QY	1342	INRFRVNAQGFYFDNNGYAVTGARTINGQHLYFRANGVQVKGEFVTDHHRISYYDG	1401	
Db	1363	YRNRF-HEIDGSWYFFDGNNAVKGMVNINGQLLFDNNGKQIKGHLVR-VNGVRYFDP	1420	
QY	1402	NSGDIQIRNFRVNAQGFYFDNNG	1426	
Db	1421	NSGEMAVNRVWVSFGWVYFDGEG	1445	
RESULT 10				
T30552				
glucosyltransferase N - Streptococcus salivarius (fragment)				
C:Species: Streptococcus salivarius				
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004				
C:Accession: T30552				
R:Jaffe, R.I.				
submitted to the EMBL Data Library, February 1998				
A:Description: Streptococcus salivarius V1477 gtfN.				
A:Reference number: Z20854				
A:Accession: T30552				
A>Status: preliminary; translated from GB/EMBL/DBJ				
A:Molecule type: DNA				
A:Residues: 1-1449 <JAF>				
A:Cross-references: UNIPROT:O68542; UNIPARC:UPI00000B10FD; EMBL:AF049609; NID:G7				
C:Genetics:				
A:Gene: gtfN				
Query Match				
Best Local Similarity 41.7%; Score 3229.5; DB 2; Length 1449;				
Matches 698; Conservative 218; Mismatches 430; Indels 179; Gaps 36;				
QY	1	MDKKVRYKLRVKRWVTVSV-----ASAVMTLTLSGLVKADSNESKS-----	45	
Db	1	MDKKVHYKWKVKQWVTVIAVGLSLGAVSAVSLGT-NDGVVQADSDHTDATVAIPDITVD	59	
QY	46	--QISNDSNTSVVTANEESNVITEATSKQEAASSQTNHT-----VTTSSSSTSVNPKVYV	99	
Db	60	TGTVSNDT-----TAAQDPTTAAATN--DVATDQATPATPDLTDTTNTVAANAVIDV	112	
QY	100	SNPYTVG-----ETASNGEKLQ--NQTITVDK--TSEAA-----	129	
Db	113	A---TVGTDRAATNTDNTATNDTAVDTTNNNTTDTTNTVAANAVIDV	169	
QY	130	--ANNISKOTTEAD---TDVIDDSNAANLQILEKLPNVKEIDGKYYYIDNNGKVRNFTL	184	
Db	170	RRATPVNGTNNANNTVTVVNDLDPATNNVVDGSPSHIKTNGKQYVVEDDGTIRKNYYL	229	

QY 185 -----IADKILHFBETGAYTDTSDTWNKDLVITRNSRLYKKNQVYDRS 229
DB 230 ERIGSQYFNAETGLSNQKRYFKNGG-TGSSADSTNTVTVNGDK-----NAPYGT 283
QY 230 AQSPEHVDHLYTAEBSWYRPKYILKDGKWTQSTEDKPRPLMTWMPDQBTQRYVNYNA 289
DB 284 DXDIELVDGYFTANTWYRPKEILKDGKWTASTENDKRPPLTVMWPSKAIQASLYNWK 343
QY 290 Q-LGINKTYDTSNQLNLNIAAATIQAKIEAKITTLKNTDMLRQTSIAFVKTSANSDS 348
DB 344 QGLGNTQYTSFSSQTMQDAALEYQKRIEGRIAREGNTDMLRTTIKNFVKTPQCNSTS 403
QY 349 EK-PFDDHLONGAVLYDNEGLKTPYANSNYRILNPTNQTGKDPRTYADNTIGYEFL 407
DB 404 ENLDNNDHLQGGALLYNDSR-TSHANSDYRLNRPSTQTKGNPKYTKOTNSNGFBL 462
QY 408 LANDVDNSNPVQAOLNHLHFMFGNIYANDPDNFDLSRVDVNDVADLQIAGDY 467
DB 463 LANDIDNSNPAQAOLNHLHYIMNIGITGSGEDNFDGVRVDVANDVNADLQIADSY 522
QY 468 LKAAGIHKNDKAANDHLSILEAWSNDPTPYLHDDGDNMINNDKRLSLLSLAKPLNQ 527
DB 523 PKAKYGADQSQDAIKHLSILEAWSNDAYNEDTKGAQLPMDPMHLALVYSLLRPIGN 582
QY 528 RSGMPLITNSLVNRTDNDNAETAAPVSYSPFRAHDSVODLIADIIKAEINPNVVGYSFT 587
DB 583 RSGVEPLISLNSDRSSESGKSKRMANYAFVRAHDSVQSIQIGIIRKNEINPQSTGNTFT 642
QY 588 MEEIKKAFEINYKOLILATEKKYTHNTALSVALLTNTKSSVPRVYVGMFTDDGQYMAHK 647
DB 643 LDMMKAFEINYKONRSANKQTQNIPIFSAVALMUTHKDTPRVYVGMFTDDGQYMAQK 702
QY 648 TINYBAIETLLKARIKYVSGQAMRNQOVGNSE-----IITSVRYKKGAKATDGD 699
DB 703 SPYDAIETLLKGRIRYAGGQDMKVNYGYGNTNGWDAAGVLTSVRYGTGANSASDTGT 762
QY 700 RTRTSQVAVTEGNPISRLKASDRVNMGAHQAQYRPLLLTTDNGIKAYHSDQAA 759
DB 763 AETRNOGMVIVSNQPALRL--TSNLTINMGAHRNQYRPLLLTTDNGVATYLNDSAN 820
QY 760 GLAVRYTNDRGELIFTAADIKGVANPQSVGLGVVPVGAALIKMPALRLARPHQOMAS-- 817
DB 821 GIVKTDGNGNLTFPANEIRGINFQDGYLAVVPVGAS--ENQDVRVAPSKENSSGL 878
QY 818 -VHQNAA LDSRMVPEGFNFQAFATKBEYTNVVIKNDVKPFAEAGVTDTFEMAPQYVST 876
DB 879 VYESNAALDSQVIYEGSFNFQDFVNPSPQYTNKKIAENANLPSKSGITSPFAPQYVSSD 938
QY 877 DGSFLDSVIQNGYATDRYDGLIGSKPKYGTADDLVKAIKALHSKGIKVMADWVDPQMYA 936
DB 939 DGSFLDSVIQNGYATDRYDGLIGSKPKYGTADDLVKAIKALHSKGIKVMADWVDPQY 998
QY 937 FPEKEVTTATRVKDYGTVPAGSOIKNTLYVVDGKSSGQQAQYKGAFLLELQAKYPELF 996
DB 999 LPGBEVTTATRVNNGYGETDGAIDHLSIYAAKTRTFGNDYQKYGGAFLDELKRLYPOIF 1058
QY 997 ARKOISTGVPMDPSVKIKQWSAKYPNGTNIILGRGAGYVLKQDQATNTYFNISDNKEINFLP 1056
DB 1059 DRVQISTGKRWTTDEKTIKWSAKYMNNGTNIILDRGSEYVLKN-GLNGYYGTNGGKV--SLP 1115
QY 1057 KTL-----LNQDSQVGFSDGKGYV-----YYSTSGYQANQTFISEGD-KWYTFDNN 1102
DB 1116 KWVGSNQNTGDNQNG---DGSQKPEKRLFSVRYRYNNGQYAKNAFIDKNDGNVYFDNS 1172
QY 1103 GYMTGAOSINGVNYFELSNGLQLRDAILKKNEDGTYYAYGNDGRYENGYQYFMSGVWRH 1162
DB 1173 GRMAVEKTIIDGKQIFFFLANGVQLRDGYRQN-----RR----- 1205
QY 1163 FNNGMSVGLTVIDGQVQYFDMGYQAQKGFVTTADGK-IRYFDKQSGNMVNRNFIENE 1221
DB 1206 -----GQVFFYDQNG-----VLSANGKQDPKPDNNNTSGRNQFVQIGN 1244

QY 1222 GKWLYLGEDGAAYTGSOTINGOHLYFRANGVQVKGEFVTDHHRISYYDGNSGDOIERNRP 1281
DB 1245 NWAYYDNGNGKRVIHQHONINGQELFFDNNGVQVNGRTVNE-NGAIRYDANSNGEMARNRP 1303
QY 1282 VRNAQGOWFYFDNNGYAVTGARTINGOLLYFRANGVQVKGEFVTDYGRISYYDGNSGDQ 1341
DB 1304 AEIEPGWYFNNBDGTAVKGSQNINGDLYFDQNGRQVKA-LANVDGNLRYDYVNSGEL 1362
QY 1342 IRNRPVNAQGOWFYFDNNGYAVTGARTINGOHLYFRANGVQVKGEFVTDHRGRISYYD 1401
DB 1363 YRNRFP-HEIDGSWYFYDGNNAVKGMVINGNQLLFDNNGKQIKGHLVR-VNGVVRYFDP 1420
QY 1402 NSGDOIERNRFRNAQGOWFYFDNNG 1426
DB 1421 NSGEMAVNRWVEVSPGWWVYFDGEG 1445
RESULT 11
A44811
glucosyltransferase (EC 2.4.1.1-) I - Streptococcus salivarius
C:Species: Streptococcus salivarius
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C:Accession: A44811; S22726; S28809
R:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.
J.Gen. Microbiol. 137, 2577-2593, 1991
A>Title: Molecular characterization of a cluster of at least two glucosyltransferase genes
A:Reference number: A44811; MUID:9214837; PMID:1838391
A:Accession: A44811
A:Molecule type: DNA
A:Residues: 1-1518 <GIF>
A:Cross-references: UNIPROT:Q00600; UNIPARC:UPI00000BEEF31; EMBL:Z11873; NID:G47526; PIDN:44811
A:Note: sequence extracted from NCBI backbone (NCBIN:81050, NCBIIP:81052)
C:Genetics:
C:Gene: gtfJ
C:Keywords: glycosyltransferase; hexosyltransferase
P:1307-1326/Domain: cpl repeat homology <CP4>
Query Match 41.3%; Score 3196.5; DB 2; Length 1518;
Best Local Similarity 45.2%; Pred. No. 6.7e-147;
Matches 695; Conservative 233; Mismatches 472; Indels 139; Gaps 31;
QY 1 MDKVRVYKLRKVRKRVTVSVASAVMTLTLLSGGL-----VKADSNESKQISNDSNTS 54
DB 1 MENIKYKHLKVKQWTVIAVAS--VALATVGLGSLVTTSSVSADETQDKVTQNSGTT 58
QY 55 -----VVTANEESNVITEATSKQEAASSQTNHVTVTSSSTS 91
DB 59 ASLVTSPKATKADKRTNTKEADVLTPAKETNAVETATTTNTQTAAEAATTATTADVA 118
QY 92 VNPKEVSNPYTVGETASNGEKLQNTTVDKTSEAAAANNISKOTTETADTVIDDSNA 151
DB 119 AVPNKEAV---VTTDAPATVTEKABEQPATV---KAEVVNTEVKAPEA---ALKDS--- 165
QY 152 NLOILEKLPNVEIDGKYVYDNNNGKVRNTFTLIADGKILHFDGTGATVDTSIDTVNKOI 211
DB 166 EVEAALSILKNIKIDGKYVYVNEGSHKENFAITVNGQLLYFGKDCALTSSTYSFTPG- 224
QY 212 VTTRSNLYKKNQVYDRSAQSFHVHDHYLTAEWSYRPKYIILKDGKWTQSTQSTEKDPRPLM 271
DB 225 TTNIVDGFSINNRAYDSSEASFELIDGYLTADSWYRPAIIKDGVTWQASTAEDFRPLM 284
QY 272 TWDPDQETQROYVNMNAQLGINKTYDTSNQLNLNIAAATIQAKIEAKITTLKNTDMLR 331
DB 285 AWPENVDVTQVNYLNYMSKVFNLDAKYSTDKQETLKVAAKDIQIKIEQIQAEKSTQMLR 344
QY 332 QTIISAFVKTSQAWNSDSEK---PFDDHLONGAVLYDNEGLKTPYANSNYRILNPTNQ 387
DB 345 ETISAFVKTPQWKNKETENYSKGGEDHLQGGALLYVNDNR-TPWANSDYRLNRTATNQ 403
QY 388 TG-----KQDPRTYADNTIGYEFLANDVNSNPVQAOLNHLFLMNFNFIYA 438
DB 404 TGTIDKSIDLQSDP-----NHMGGFDFLLANDVDSLNPVQAOLNQLIHYLMNWSIWM 458

QY 921 -----KGIKVMADWV-----PDMYAPPEKEVVVATRV-----D 949
Db 1020 GYIYFDSTGCKAQKGFVTLGKTYTNTNTMTYFVNANNNLYYFDNEGVMQTWINYN 1079
QY 950 KY-----GTPVAGSQIKNTLYVVDGKSSGKQOQAKYGGAPLELQAKYPFLPARKOISTG 1004
Db 1080 RYFSGATGASVTGQ-----TIDGNKYCFDSNG-----AIYTDVVTINGSTYG 1122
QY 1005 VPMPPSVKIKOMSAKYFNGTNIILGRGAYVYLKQDQATNTYFNISD-----NKEIN 1053
Db 1123 FNTD-GIMLTGMQTIYRN-----RGY-----SSYFNTYFNSDGTAKTGPFYLNKTY 1170
QY 1054 FLPKTLNODSQV-----GFSVDGKGYVYSTSGYQAKNTFISEGDKWYFYFNNNGVMYTGAS 1111
Db 1171 FNP-----SDGRMLQYQYVINGNHYYFAPDG-TMQTGWITNSSSKYLYDPGSAVATGLQT 1224
QY 1112 INGVNYFELNSGLDAILKNEGTG---AYGNDGRYENGYYQYFMSGVHRHFNNGEM 1168
Db 1225 INGNKYCFDSNG-----ILQHNIGFYIGNTYGSD-----NNGIM 1259
QY 1169 SVGLTVIDGOVQYFDEMGOYQAKGFVTTADGKIRYFDKQS-----GNMYNRNFIENBEGK- 1223
Db 1260 LGLQLQLINGLYLFCNSDGSVKTG--LVTYLGTKYTFDSYVSGFQNNNTYFNGDGTM 1317
QY 1224 ----WL-----YLGEDGAATGSGQTINGQHLRYFRANGVQVKGFEVTDHHRISYYDGS 1273
Db 1318 QTGWNYGYRYLYLNDSGIKVTGWOTIDGNKYF-----DYYGAKT 1358
QY 1274 GQDIRNFRVNAQGWYFEDNNGYAVTGARTINGQLLYFRANGVQVKGFEVTT----- 1325
Db 1359 G-----IVNDIGNTYFNNSGVMLTGWOHINGSTYFNSGIANWG-FITYLGTKYTF 1410
QY 1326 DRYGRI-----SYDNGSG-----DQIRNFRVNA----- 1350
Db 1411 DSYGRWQISMTINGTSYFYANGVMKSTSDPNTLAVCWVDSYFYQYLYLNAAGTKLGT 1470
QY 1351 ----OQWYFDPNNGYAVTGARTINGQHLRYFRANGVQVKG-----E 1387
Db 1471 LOTIDGNTYFDSNGIMQTGIITINGNRYGFGVGVMLYGLQFINNNTYNSYGISQSG 1530
QY 1388 FVT-----DRHGR-----SYDNGS-----GDOI 1407
Db 1531 FVTLGNTYFDSYGEIRGLTYINNNYYFNSKGMETGWSYLYRYANPNGLITLGTGT 1590
QY 1408 RN--RFVRNAQGWYF-----FDNNGYAVTGARTINGQHLRYFRANGVQVKGFEV 1454
Db 1591 INKTYFNSDGLLYDLQYNGSYGFDKXGVMLYGLQITGNTYLYNSGISQSG-FI 1649
QY 1455 TDYGRISYYDANSGERVRI 1474
Db 1650 T-LNGKTYFDSYSGWRTGI 1668

RESULT 14
A37184
glucan-binding protein - Streptococcus mutans
C:Species: Streptococcus mutans
C>Date: 31-Jan-1992 #sequence_revision 31-Jan-1992 #text_change 09-Jul-2004
C:Accession: A37184
R:Banas, J.A.; Russell, R.R.B.; Ferretti, J.J.
Infected. Immun. 58, 667-673, 1990
A:Title: Sequence analysis of the gene for the glucan-binding protein of Streptococcus m
A:Reference number: A37184; MUID:90170123; PMID:2307516
A:Accession: A37184
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-563 <BAN>
A:Cross-references: UNIPARC:UPI000008CB5; GB:M30945; NID:G153637; PIDN:
F:169-188/Domain: cpl repeat homology <CP1>
F:264-283/Domain: cpl repeat homology <CP2>
F:349-368/Domain: cpl repeat homology <CP3>
F:504-523/Domain: cpl repeat homology <CP4>
F:525-548/Domain: cpl repeat homology <CP5>

Query Match 6.1%; Score 473; DB 2; Length 563;
Best Local Similarity 30.7%; Pred. No. 15e-15;
Matches 139; Conservative 75; Mismatches 145; Indels 94; Gaps 22;
QY 1106 VTG-AQSINGVNYFELNSGLDAILKNEGTGTYAYYNGDGRYENGYYQYFMSGVHRHFN 1164
Db 122 VSGEAQSVNAP---SENAAQETA--KTEPATAA--ENNDAAPTNSFFK-KDGKWKYK 173
QY 1165 -NGEMSVGLTVIDGOVQYFDEMGOYQAKGF-VTTAD-----GKIRYF 1204
Db 174 ADGQLATGWQIIDGKQLYFNQDSQVKGIEHVETGQIYHPVFISDPSVLEWNKIYF 233
QY 1205 DRQSGMNYNRFI-----EN--BEKWLILGEDGAATVTSOTINGQHLRYFRANGV 1252
Db 234 DPDSGELWKDRFVYSYADPLHYENIKHEG-WFYLGEDGKAALGWRTIGKGYFDTNGV 292
QY 1253 QVKG-----FVTDHHRISYYDNGSGDQIRNFRV-----NAQG-----QWYF 1292
Db 293 QVKGKLISTDGNVNLISQYKKSFLDPDTGEATNRFVNAKYFYFNAGYVSTTDWFM 352
QY 1293 DNNGYAVTGARTINGQLLYFR-ANGVQVKGFEVTDYGRISYYDNGSGDQIRNFR----- 1346
Db 353 GADGIGVTDWQKIDGMHYFFESPSSGIVQKGD-IAERDGGYVYLDSDSGVQVKNRFTTTPA 411
QY 1347 -----VRNAQGWYFEDNNGYAVTGARTINGQHLRYF-----RANGVQVKGFEVTD 1391
Db 412 ERISTVEARFPKTYFAGDSRKDLTGWQIIDGKTYFKDDHSIKAKSYSGISGVPDD 471
QY 1392 RHGRI---SYDNGSGDQIRNFRV--NAQGWYFDPNNGYAVTGARTINGQHLRYF----- 1442
Db 472 GFAIDGCGYFPDTCQGFVTRNFRVKYDYSNTWYVYSGDKRVSGWQTIDGKRYYSQDE 531
QY 1443 RANGVQVKGFEVTDYGRISYYDANSGERVRI 1475
Db 532 KTKGQIRGQITID-GREYTFDKDSGEVINSN 563

RESULT 15
A37052
toxin A - Clostridium difficile
C:Species: Clostridium difficile
C>Date: 31-Jan-1992 #sequence_revision 31-Jan-1992 #text_change 24-Sep-1999
C:Accession: A37052; A60991; S21897; S22437; S08638
R:Dove, C.H.; Wang, S.Z.; Price, S.B.; Phelps, C.J.; Lyster, D.M.; Wilkins, T.D.; Johns
Infected. Immun. 58, 480-488, 1990
A:Title: Molecular characterization of the Clostridium difficile toxin A gene.
A:Reference number: A37052; MUID:90129305; PMID:2105276
A:Accession: A37052
A:Molecule type: DNA
A:Residues: 1-2710 <DOV>
A:Cross-references: UNIPARC:UPI000016EAD4; GB:M30307; NID:G144925; PIDN:AAA23283.1; PID
R:Wren, B.W.; Clayton, C.B.; Tabagchali, S.
FEMS Microbiol. Lett. 70, 1-6, 1990
A:Title: Nucleotide sequence of Clostridium difficile toxin A gene fragment and detecti
A:Reference number: A60991
A:Accession: A60991
A:Molecule type: DNA
A:Residues: 1-154 <EIC>
A:Cross-references: UNIPARC:UPI0000178591; EMBL:X60984
R:von Eichel-Streiber, C.; Laufenberg-Feldmann, R.; Sartingen, S.; Schulze, J.; Sauerbor
Mol. Gen. Genet. 233, 260-268, 1992
A:Title: Comparative sequence analysis of the Clostridium difficile toxins A and B.
A:Reference number: S22434; MUID:92293124; PMID:1603068
A:Accession: S22437

Db 2142 GVTIINGKHFFNTDGIQIGVF--KGPNGFEYFAPANTDANNIEQQAILYQNEFL-TLN 2198
QY 1222 GKWLYIGEDGAAVTGSQTINGQHLYFRAN-----GVOVKGEFVT 1260
Db 2199 GKYYFGSDSKAVTGWRIINNKYYFNPNAIAAHLCTINNDKYFSDGILQNGYITI 2258
QY 1261 DHGRISYYDGNSGDOI-----RNRFRNAQQOM 1289
Db 2259 ERNN--FYFDANNESKMTVTGPKPNGFEYFAPANTHNNNIEQAIIVYQNKEL-TLNGKK 2315
QY 1290 FYFDNNGVAVTGARTINGQLLYFRANGVQVKGEFVTDYGRISYYDGNSGDOI RNRFRN 1349
Db 2316 YIFDNDKAVTGWQIDGKKYFNLTAEATGWQID-GKKYFNLTAEATG--WOT 2372
QY 1350 AQQOMFYFDNNGY-AVTGARTINGQHLYFRANGVQVKGEFVTDYGRISYY-----DGN 1403
Db 2373 IDGKKYFNTNTFIASGTYSINGKHFFYNTDGIQIGVF--KGPNGFEYFAPANTDANN 2430
QY 1404 --GDOI--RNRFRNAQQOMFYFDNNGVAVTGARTINGQHLYFRAN-GVOVKGEFVTDY 1458
Db 2431 IEGQAILYQNKEL-TLNGKKYFGSDSKAVTGLRTIDGKKYFNTNTAVAVTGWQIN-- 2487
QY 1459 GRISYYDANS 1468
Db 2488 GKYYFNTNT 2497

Search completed: February 11, 2006, 19:41:45
Job time : 40.4446 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 18:57:35 ; Search time 199.028 Seconds
(without alignments)
5228.676 Million cell updates/sec

Title: US-10-797-821-34
Perfect score: 7741
Sequence: 1 MDKVRVYKLRVKRVTVS.....DRYGRISYDANSRVRIN 1475

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_05.80.*

1: uniprot_eprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7535.5	97.3	1476	1	GTFB_STRMU
2	5551.5	71.7	1455	1	GTFB_STRMU
3	4582	59.2	1590	2	Q55263_9STRE
4	4580	59.2	1590	2	Q59983_9STRE
5	4578.5	59.1	1597	1	GTF1_STRDO
6	4531	58.5	1592	1	GTF2_STRDO
7	3879.5	50.1	1462	1	GTFD_STRMU
8	3669.5	47.4	1577	2	Q54178_STRGN
9	3650	47.2	1454	2	Q69A94_LEUME
10	3597.5	46.5	1575	2	Q9LCH3_STROR
11	3523.5	45.5	1512	2	Q9WXJ5_9STRE
12	3512.5	45.4	1506	2	Q56CX8_9STRE
13	3506	45.3	1577	2	Q55265_STRSL
14	3434.5	44.4	1554	2	Q8XZL5_9STRE
15	3398.5	43.9	2835	2	Q8G9Q2_LEUME
16	3289	42.5	1599	2	Q00599_STRSL
17	3246	41.9	1290	2	Q48756_LEUME
18	3243	41.9	1508	2	Q52224_LEUME
19	3238.5	41.8	1449	2	Q52264_STRSL
20	3230	41.7	1508	2	Q9EZH5_LEUME
21	3229.5	41.7	1449	2	Q68542_STRSL
22	3215.5	41.5	1477	2	Q9L466_LEUME
23	3197	41.3	1527	2	Q8KRE1_LEUME
24	3196.5	41.3	1518	2	Q00600_STRSL
25	3178.5	41.1	1330	2	Q84CN4_LEUME
26	3171.5	41.0	1527	2	Q9ZAR4_LEUME
27	3168.5	40.9	1522	2	Q6TXV4_LEUME
28	3160	40.8	1561	2	Q5SBM8_9LACO
29	2945.5	38.1	1365	1	GTF5_STRDO
30	2897	37.4	1338	2	Q9WXJ4_9STRE
31	2877.5	37.2	1595	2	Q5SBM3_LACSK

32	2793	36.1	1463	2	Q5SBM6_LACFE
33	2537	32.8	2057	2	Q9RE05_LEUME
34	2402	31.0	1016	2	Q9LCJ7_LEUME
35	2343	30.3	1781	2	Q4JCS4_LACRE
36	2340.5	30.2	1772	2	Q5SBN0_LACRE
37	2340	30.2	1772	2	Q5SBN3_LACRE
38	2234	30.2	1781	2	Q4JLC7_LACRE
39	2280	29.5	1781	2	Q5SBL9_LACRE
40	2272	29.4	591	2	Q8VUH3_STRMU
41	1507.5	19.5	1619	2	Q5SBM0_LACRE
42	1493.5	19.3	1231	2	Q5SEN1_LACRE
43	930	12.0	522	2	Q8VV10_STRSA
44	794	10.3	374	2	Q6ZX19_9LACO
45	553	7.1	2817	2	Q97K42_CLOAB

RESULT 1

GTFB_STRMU STANDARD; PRT: 1476 AA.

AC P08987; Q69381; Q69384; Q69387; Q69390; Q69396;

DT 01-NOV-1988 (Rel. 09, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 10-MAY-2005 (Rel. 47, Last annotation update)

DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranase)

DE (Sucrose 6-glucosyltransferase).

GN Name=gtfB; OrderedLocNames=SMU.1004;

OS Streptococcus mutans.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI_TaxID=1309;

RN [1]_

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=GS-5;

RX MEDLINE=87308013; PubMed=3040685;

RA Shiroza T., Ueda S., Kuramitsu H.K.;

RT "Sequence analysis of the gtfB gene from Streptococcus mutans.";

RL J. Bacteriol. 169:4263-4270(1987).

RN [2]_

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=MT4239 / Serotype C, MT4245 / Serotype e, MT4251 / Serotype f, MT4467 / Serotype e, and MT8148 / Serotype c;

RX MEDLINE=98231643; PubMed=9570124;

RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S., Kimura S., Hamada S.;

RT "Molecular analyses of glucosyltransferase genes among strains of Streptococcus mutans.";

RL FEMS Microbiol. Lett. 161:331-336(1998).

RN [3]_

RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

RC STRAIN=UAI59 / ATCC 700610 / Serotype c;

RX MEDLINE=2295063; PubMed=12397186; DOI=10.1073/pnas.172501299;

RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J., Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P., Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A., Ferretti J.J.;

RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental pathogen.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).

CC -!- FUNCTION: Production of extracellular glucans, that are thought to play a key role in the development of the dental plaque because of their ability to adhere to smooth surfaces and mediate the aggregation of bacterial cells and food debris.

CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-fructose + (1,6-alpha-D-glucosyl) (n+1).

CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both forms of glucans.

CC -!- SIMILARITY: Belongs to the glucosyl hydrolase 70 family.

Q5sbm6 lactobacill
Q9re05 leuconostoc
Q9lcj7 leuconostoc
Q4jcs4 lactobacill
Q5sbm0 lactobacill
Q5sbm3 lactobacill
Q4jlc7 lactobacill
Q5sbl9 lactobacill
Q8vuh3 streptococ
Q5sbm0 lactobacill
Q5sen1 lactobacill
Q8vv10 streptococ
Q6zx19 lactobacill
Q97k42 clostridium

```

CC CC      -!- SIMILARITY: Contains 10 cell wall binding repeats.
CC CC      -----
CC CC      This Swiss-Prot entry is copyright. It is produced through a collaboration
CC CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC      the European Bioinformatics Institute. There are no restrictions on its
CC CC      use as long as its content is in no way modified and this statement is not
CC CC      removed.
CC CC      -----
DR DR      EMBL; M17361; AAA88588.1; -; Genomic DNA.
DR DR      EMBL; D88651; BAA26101.1; -; Genomic DNA.
DR DR      EMBL; D88654; BAA26105.1; -; Genomic DNA.
DR DR      EMBL; D88657; BAA26109.1; -; Genomic DNA.
DR DR      EMBL; D88660; BAA26113.1; -; Genomic DNA.
DR DR      EMBL; D89977; BAA26119.1; -; Genomic DNA.
DR DR      EMBL; AE014940; AAN58705.1; -; Genomic DNA.
DR DR      PIR; B33135; B33135.
DR DR      HSGP; P06653; 1H8G.
DR DR      InterPro; IPR002479; Cell wall bd put.
DR DR      InterPro; IPR003318; Glyco_hydro_70.
DR DR      Pfam; PF01473; CW_binding_1; 4.
DR DR      Pfam; PF02324; Glyco_hydro_70; 1.
KW KW      Complete proteome; Dental caries; Glycosyltransferase; Repeat; Signal;
KW KW      Transferase.
FT FT      SIGNAL      1      34      Potential.
FT FT      CHAIN      35      1476      Glucosyltransferase-I.
FT FT      REPEAT      1097      1130      A repeat.
FT FT      REPEAT      1161      1210      1.
FT FT      REPEAT      1225      1275      2.
FT FT      REPEAT      1290      1340      3.
FT FT      REPEAT      1355      1405      4.
FT FT      REPEAT      1420      1470      5.
FT FT      REGION      35      1051      Catalytic (approximate).
FT FT      REGION      1097      1476      Glucan-binding (approximate).
FT FT      REGION      1161      1470      5 X tandem repeats.
FT FT      VARIANT      62      62      S -> T (in strain MT4239).
FT FT      VARIANT      65      65      T -> I (in strain GS-5).
FT FT      VARIANT      68      68      V -> A (in strain GS-5, strain MT4245,
FT FT      strain MT4251, strain MT4467 and strain
FT FT      MT8148).
FT FT      Q -> P (in strain MT4251).
FT FT      I -> S (in strain GS-5, strain MT4245,
FT FT      strain MT4251, strain MT4467 and strain
FT FT      MT8148).
FT FT      S -> F (in strain MT4251).
FT FT      K -> N (in strain GS-5, strain MT4467 and
FT FT      strain MT8148).
FT FT      S -> D (in strain GS-5, strain MT4467 and
FT FT      strain MT8148).
FT FT      N -> R (in strain MT4239).
FT FT      I -> T (in strain MT4239).
FT FT      K -> R (in strain MT8148).
FT FT      F -> Y (in strain MT8148).
FT FT      T -> I (in strain MT8148).
FT FT      A -> V (in strain MT8148).
FT FT      F -> L (in strain MT8148).
FT FT      FGKPE -> YGTPVA (in strain GS-5, strain
FT FT      MT4239 and strain MT4467).
FT FT      SV -> NT (in strain GS-5, strain MT4239
FT FT      and strain MT4467).
FT FT      ADS -> VDG (in strain GS-5, strain MT4239
FT FT      and strain MT4467).
FT FT      A -> T (in strain MT4239).
FT FT      S -> N (in strain MT4239).
FT FT      H -> Y (in strain MT4251).
FT FT      E -> K (in strain MT8148).
FT FT      Y -> C (in strain MT8148).
FT FT      A -> P (in strain MT4239).
FT FT      R -> H (in strain GS-5 and strain
FT FT      MT4467).
FT FT      R -> P (in strain MT8148).
FT FT      Y -> H (in strain GS-5, strain MT4239,
FT FT      strain MT4467 and strain MT8148).
FT FT      S -> G (in strain GS-5 strain MT4239,
FT FT      strain MT4467 and strain MT8148).

```

```

FT VARIANT      1329      1329      H -> Y (in strain GS-5 and strain
FT MT4467).
FT VARIANT      1394      1394      Y -> H (in strain GS-5, strain MT4239,
FT strain MT4467 and strain MT8148).
FT VARIANT      1402      1402      S -> G (in strain GS-5, strain MT4239,
FT strain MT4467 and strain MT8148).
FT VARIANT      1459      1459      Y -> H (in strain MT4467).
FT CONFLICT      570      570      R -> A (in Ref. 1).
FT CONFLICT      800      817      ADQDVRVAASAPSTDGK -> LIKMPALRLARPHQOMA
FT (in Ref. 1).
FT CONFLICT      1310      1310      H -> L (in Ref. 1).
FT SEQUENCE      1476      AA; 165847 MW; 9C6B09F731B4CBCF CRC64;

Query Match      97.3%; Score 7535.5; DB 1; Length 1476;
Best Local Similarity 97.6%; Pred. No. 0;
Matches 1442; Conservative 8; Mismatches 23; Indels 5; Gaps 2;

QY      1      MDKKVRYKLRVKRKRWTVSVASAVMTLTTLSSGGLVKAADSNEKSQISNDSNTSVVTANE 60
DB      1      MDKKVRYKLRVKRKRWTVSVASAVMTLTTLSSGGLVKAADSNEKSQISNDSNTSVVTANE 60
QY      61      ESNVITEATSKQEAASSQTNHTVTTSSTSSSTSVVNPKEVSNPYTVGTASNGEKLQNTT 120
DB      61      ESNVTEVTSKQEAASSQTNHTVTTSSTSSSTSVVNPKEVSNPYTVGTASNGEKLQNTT 120
QY      121      TVDKTSEAAANNISKQTEADTVDDSNAAANLQILEKLPNVKEIDGKYIYDNNKQVRT 180
DB      121      TVDKTSEAAANNISKQTEADTVDDSNAAANLQILEKLPNVKEIDGKYIYDNNKQVRT 180
QY      181      NFTLIADGKILHFDGTGAYTDTSIDTVNKNIDVITRSNLYKKYNQVYDSASQSFHVHDVYL 240
DB      181      NFTLIADGKILHFDGTGAYTDTSIDTVNKNIDVITRSNLYKKYNQVYDSASQSFHVHDVYL 240
QY      241      TAESWYRPKYILKDGKTWTQSTEKDFRPLMTWPFQETQRQYVYNNMAQLGINKTYDDT 300
DB      241      TAESWYRPKYILKDGKTWTQSTEKDFRPLMTWPFQETQRQYVYNNMAQLGINKTYDDT 300
QY      301      SNQLNLNIAAATIOAKIEAKITTLKNTDWLRQTISAFVKTSAMNSDSEKFPDDHLONGA 360
DB      301      SNQLNLNIAAATIOAKIEAKITTLKNTDWLRQTISAFVKTSAMNSDSEKFPDDHLONGA 360
QY      361      VLYDNEGKLTYPYANSNYRILNRTPTNQTKKDPRTADNTTGGVEFLANDVDNSNPVQ 420
DB      361      VLYDNEGKLTYPYANSNYRILNRTPTNQTKKDPRTADNTTGGVEFLANDVDNSNPVQ 420
QY      421      AEQLNLWHLFMNFGNIYANDPDANFDSIRVDAVNDVADLLQIAGDYLKAAKGIHKNDKA 480
DB      421      AEQLNLWHLFMNFGNIYANDPDANFDSIRVDAVNDVADLLQIAGDYLKAAKGIHKNDKA 480
QY      481      ANDHLSILEAWSNDPTPYLHDDGDNMINMDNKLRLSLFLSLAKPLNQSGMNPILTNSLV 540
DB      481      ANDHLSILEAWSNDPTPYLHDDGDNMINMDNKLRLSLFLSLAKPLNQSGMNPILTNSLV 540
QY      541      NRTDNEAETAAPPSYFTRAHDSQVODLIADIKAEINPNVVGVSFTWEEIKKAEIYNK 600
DB      541      NRTDNEAETAAPPSYFTRAHDSQVODLIADIKAEINPNVVGVSFTWEEIKKAEIYNK 600
QY      601      DLLATEKKYTHYNTALSYALLLTNKSVPVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 660
DB      601      DLLATEKKYTHYNTALSYALLLTNKSVPVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 660
QY      661      RIKYVSGQAMRNQOVGNSEIITSVRYKQKALKATDTGDRTRTSVGVAVIEGNNPSLRUK 720
DB      661      RIKYVSGQAMRNQOVGNSEIITSVRYKQKALKATDTGDRTRTSVGVAVIEGNNPSLRUK 720
QY      721      ASDRVVNMGAHKNQAVRPLLLTTDNGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKG 780
DB      721      ASDRVVNMGAHKNQAVRPLLLTTDNGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKG 780
QY      781      YANPQVSGYLGVVWVVGAAALIKMFALRLA---RPHQOQMAVHQNAALDSRVNFEFSNFQ 837
DB      781      YANPQVSGYLGVVWVVGAAADQ--DVRVAASTAPSTDGKSVHQNAALDSRVNFEFSNFQ 838

```


FT	VARIANT	150	151	SR -> PK (in strain GS-5, strain MT4239 and strain MT4467).
FT	VARIANT	256	256	A -> V (in strain GS-5 and strain MT4467).
FT	VARIANT	425	425	R -> N (in strain MT4251).
FT	VARIANT	519	519	Y -> D (in strain MT4245 and strain MT4251).
FT	VARIANT	538	538	R -> K (in strain MT4245 and strain MT4251).
FT	VARIANT	549	545	Y -> F (in strain MT4245 and strain MT4251).
FT	VARIANT	597	597	N -> D (in strain MT4245, strain MT4251, strain MT4467 and strain MT8148).
FT	VARIANT	600	600	R -> K (in strain MT4245, strain MT4251, strain MT4467 and strain MT8148).
FT	VARIANT	601	601	A -> T (in strain GS-5).
FT	VARIANT	614	614	M -> T (in strain GS-5).
FT	VARIANT	727	727	T -> I (in strain MT8148).
FT	VARIANT	734	734	A -> V (in strain MT8148).
FT	VARIANT	964	964	L -> F (in strain MT4239).
FT	VARIANT	1113	1113	N -> Y (in strain MT4239).
FT	VARIANT	1118	1118	A -> T (in strain MT4239).
FT	VARIANT	1204	1204	I -> V (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).
FT	VARIANT	1208	1208	V -> I (in strain MT8148).
FT	VARIANT	1292	1294	DGH -> NGY (in strain GS-5, strain MT4467 and strain MT8148).
FT	VARIANT	1305	1369	Missing (in strain MT4245).
FT	VARIANT	1326	1326	I -> V (in strain MT8148).
FT	VARIANT	1331	1331	T -> A (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).
FT	VARIANT	1377	1377	R -> K (in strain MT8148).
FT	VARIANT	1398	1398	V -> I (in strain MT8148).
FT	VARIANT	1424	1424	D -> I (in strain MT4239).
FT	VARIANT	1439	1439	V -> I (in strain MT4239 and strain MT8148).
FT	VARIANT	1444	1444	S -> P (in strain MT8148).
FT	CONFLICT	1337	1455	ORLYPKSGVQAKGELITERKGRIKYYPDNGSNEVRNRYR TSSGNYFFGNDGYALIGHVVEGRRVYFDENGVRYASHD ORNHWDYRDRDFGRGSSAVRFHRSRNGFFDNFRF -> HASILMLVRLRESSLQSVKVSNTMILLPEMKFVIM (in Ref. 1).
SQ	SEQUENCE	1455 AA;	162966 MW;	3CB455A99A4FEC86 CRC64;
Query Match				
Best Local Similarity 71.7%; Score 5551.5; DB 1; Length 1455;				
Matches 1086; Conservative 112; Mismatches 201; Indels 77; Gaps 16;				
Qy	1	MDKKRYKLRKVKKGVTVSVASAVMTLTLSGGLVKADSNESKSOISNDSNTSVVTANE	60	
Db	1	MEKKRVFLRKVKKGVTVSVASAVVTLTSLGSLVKADSTDDROQAVTESQASLVTTSE	60	
Qy	61	-----ESNVITEATS KOEAAASQTNHTVTTSSSTSVVNPKEVSNPYTVGETASN	111	
Db	61	AAKETLTATDSTATSATSQPTA--TVTDNVSTTNQSTNTTANTANFDVKPTTSE----	114	
Qy	112	GEKLNQTTTVDK-----TSEA-----AANNISKQTTAEAD-----	141	
Db	115	-----QSKTDNSDKIIATSKAVNRLTATGKFVPANNNTAHSRTVTDKVIPKPKICKLQKP	170	
Qy	142	TDVIDDSNAANLOILEKLPNVEIDGKYYYDNNCKVRTNFTLIADGKILHFDETGATD	201	
Db	171	SSLSQDDIAA-----LGNVKNIRKVNKGKYYYKEDGTLOKNKVALNINGKTFPFDGTGALSN	226	
Qy	202	TSIDTVNKD IVTT--RSNLVKKNQVYDSRQSFHVHDVHLYTAESWYRPKYILKDKGKTWTQ	260	
Db	227	NTLPSKKNITNDNTNFAQYNQVYSTDAANFEHVDHLYTAESWYRPKYILKDKGKTWTQ	286	
Qy	261	STEKDFRLLMTWPDQRTQRYVNMNAQLGINKTYDDTSNQLQNLAAATIAQAKIEAK	320	
Db	287	STEKDFRLLMTWPDQRTQRYVNMNAQLGIHQHTYNTATSPQLNLAAQTIQTKIEEK	346	
Qy	321	ITTLKNTDNLQRTISAFVKVTSQAMNSDSEKFPDDHLQNGAVLYDNEGKLTPYANSNYRIL	380	

Db	347	ITAEKNTNWLQRTISAFVKVTSQAMNSDSEKFPDDHLQKGALLYSNNSKLTQANSNYRIL	406	
Qy	381	NRTPTNQTGKDDPRYTADNTTGGYEFLANDVDSNPVVQAEQLNWLHFLMNFIGNIYAND	440	
Db	407	NRTPTNQTGKDDPRYTADRTTGGYEFLANDVDSNPVVQAEQLNWLHFLMNFIGNIYAND	466	
Qy	441	PDANPDSIRVDVADVNDADLLQIAGDYJKAAGIHKNDKAANDHLSILEASNDNDTPYLH	500	
Db	467	PDANPDSIRVDVADVNDADLLQIAGDYJKAAGIHKNDKAANDHLSILEAWSYNDTPYLH	526	
Qy	501	DGDNMINMDNKLRLSLFLSLAKPLNORSGNPNLITNSLVNRTDDNAETAAPVPSYFIRA	560	
Db	527	DGDNMINMDNKLRLSLFLSLAKPLNORSGNPNLITNSLVNRTDDNAETAAPVPSYFIRA	586	
Qy	561	HDSEVQDLIADI IKAIEINPNVVGYSFTMEEIKKAFEIYNKOLLATEKKYTHYNTALSVAL	620	
Db	587	HDSEVQDLIRNI IRAEINPNVVGYSFTMEEIKKAFEIYNKOLLATEKKYTHYNTALSVAL	646	
Qy	621	LLTNKSSVPRVYVYGDMTDDGOYMAHKTINTEALETLLKARIKYVSGGOAMRNQOVGNSE	680	
Db	647	LLTNKSSVPRVYVYGDMTDDGOYMAHKTINTEALETLLKARIKYVSGGOAMRNQOVGNSE	706	
Qy	681	IITSVRYGKGALKATDGTDRITRTSGVAVIEGNNPSLRLKASDRVVVNMGAHKNQAYRP	740	
Db	707	IITSVRYGKGALKATDGTDRITRTSGVAVIEGNNPSLRLKASDRVVVNMGAHKNQAYRP	766	
Qy	741	LLLTNDNGIKAYHSDQEAAGLVRYTNDRGELI FTAADIKGYANPOVSGVLGVVVPVGAAL	800	
Db	767	LLLTNDNGIKAYHSDQEAAGLVRYTNDRGELI FTAADIKGYANPOVSGVLGVVVPVGAAL	826	
Qy	801	IKMFALRLA----RPHQQMAVHQNAALDSRVMEFGFSNFQAFATKKEEYTNVVIANKVDK	857	
Db	827	DQ--DVRVAASTAPSTDGSKSVHQNAALDSRVMEFGFSNFQAFATKKEEYTNVVIANKVDK	884	
Qy	858	FAEKGVTDFENAPQVSVSTDGSLDSVLONGVAFYDRLGIGSKPNKYGTDADDLKAICA	917	
Db	885	FAEKGVTDFENAPQVSVSTDGSLDSVLONGVAFYDRLGIGSKPNKYGTDADDLKAICA	944	
Qy	918	LHSGIKVMADWPDQMYAFPEKEVVTATRVYDVTGTPVAGSQIKNTLVVDGSKSGKDOQ	977	
Db	945	LHSGIKVMADWPDQMYALPEKEVVTATRVYDVTGTPVAGSQIKNTLVVDGSKSGKDOQ	1004	
Qy	978	AKYGAFLEELQAKYPFELFARKQISTGVMPDPSVKIKOWSAKYFNGTILGRGAGYVLKD	1037	
Db	1005	AKYGAFLEELQAKYPFELFARKQISTGVMPDPSVKIKOWSAKYFNGTILGRGAGYVLKD	1064	
Qy	1038	QATNTYFN- ISDNKEINFLPKTLLN-----QDSQVGFSDGKGYVYVYSTSGYQAKNTFTS	1091	
Db	1065	QATNTYFSLVSDN---TFLPKSLVNPNGTSSSVTGLVDFGKGYVYVYSTSGYQAKNAFTS	1121	
Qy	1092	EGDKWYYPDNNGYVMVTGAQSIYNGVNYFLSNGLQLRDAILKNEDGTAYAYGNDGRRYENG	1151	
Db	1122	LGNWYYPDNNGYVMVTGAQSIYNGVNYFLSNGIQLRNAI YDNGNKVLSYNGNDGRRYENG	1181	
Qy	1152	YYQFMSGVWRHFNNGEMSVGLTVTDGQVQYFDEMGYQAKGPFVTTADKIRYFDQSGNM	1211	
Db	1182	YYLF--GQWRVYFQNGIMAVGLTRIHGAQVYFDASGFQAKGQFITTADGKRLYFDRDSGNQ	1240	
Qy	1212	YRNFIEEKGKWLVLGEDGAAVTGSQTINGOHLVFRANGVOVKGEFVTDHHRISYYDG	1271	
Db	1241	ISNRPVRNRSKGEWFLFDHNGVAVTGVTFNGQRLYFKPENGVOAKGEFTRDADGHLRYDDP	1300	
Qy	1272	NSGQDIRNRFVRNAQGWFFYFDNNGYAVTGARTINGQLLYFRANGVOVKGEFVTDYGR	1331	
Db	1301	NSGNEVRNRFVRNRSKGEWFLFDHNGIATVGTTRVNGQRLYFKPSNGVQAKGELLTERKGR	1360	
Qy	1332	SYDNGSDQDIRNRFVRNAQGWFFYFDNNGYAVTGARTINGOHLVFRANGVOVKGEFVTD	1391	
Db	1361	KYDPSNGNEVRNRYVRTSSGNWYVFGNDGVALIGWHVVEGRRVYFDENGVRVYASHDOR	1420	
Qy	1392	RHGRISY----YDNGSGDQDIRNRFVRNAQGWFFYFDN	1424	

Dd	1421	NHWDYDRDRFCRGSSSAVRFRHSRNG-----FFDN 1451
RESULT 3		
O55263_9STRE PRELIMINARY;	PRT; 1590 AA..	
AC O55263_9STRE PRELIMINARY;		
DT 01-NOV-1996 (TrEMBLrel. 01, Created)		
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)		
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)		
DE GTF-I.		
GN Name=Glucosyltransferase;		
OS Streptococcus sobrinus.		
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;		
CC Streptococcus.		
OX NCBI_TaxID=1310;		
RN [1]		
RP NUCLEOTIDE SEQUENCE.		
RC STRAIN=ATCC 33478;		
RA Sato S.;		
RT "DNA and amino acid sequences of water-insoluble-glucan synthetase		
RT produced from Streptococcus sobrinus ATCC 33478.";		
RL Kagoshima Daigaku Shigakubu Kiyo 16.23-29(1996).		
RN [2]		
RP PROTEIN SEQUENCE.		
RX MEDLINE=91224988; PubMed=1827439;		
RA Mooser G., Hefta S.A., Paxton R.J., Shively J.E., Lee T.D.;		
RT "Isolation and sequence of an active-site peptide containing a		
RT catalytic aspartic acid from two Streptococcus sobrinus alpha-		
RT glucosyltransferases.";		
RL J. Biol. Chem. 266:8916-8922(1991).		
DR EMBL; D63570; BAA09792.1; -; Genomic_DNA.		
PR PIR; A39841; A39841.		
DR HSP; P06653; 1GVM.		
GO GO:0009250; P:glucan biosynthesis; IEA.		
DR InterPro; IPR002479; CW binding.		
DR InterPro; IPR003318; Glyco_hydro_70.		
DR Pfam; PF01473; CW binding_1; 3.		
DR Pfam; PF02324; Glyco hydro 70; 1.		
SQ SEQUENCE 1590 AA; 176058 MW; 9DF7A3F2C6E4FD43 CRC64;		
Query Match		
Best Local Similarity 56.5%; Pred. No. 6.2e-205;		
Matches 898; Conservative 212; Mismatches 336; Indels 144; Gaps 19;		
Qy 1 MDKVRVKLRKKRWTVSVASAVMTLTTLTSLGLVKADSNESKQISNDNTSVVTANE 60		
Dd 1 MEKVRFAMKVKRWTLVSASATMLASALGHASVASADTTA-----SDSNAQAVVTGDQ 56		
Qy 61 ESNVITEATSKQEAASSQTNHVTITSSSSSTSVPNPKE-----VVSNPYTVGETASNGEKL- 115		
Dd 57 TTN--NQATDQTSIAATSEQSASTDAATDAQASAASQTGTASTDTTAQTITTANEAK 114		
Qy 116 -----NQQTVDTKTSEAANNISKOTTETADTDVIDDSNAANQLILEKLPNVKEIDGKY 169		
Dd 115 WVPTEENQGFTDEMLAE--AKOVA--TAESDSFPSPD-----LAKMSNVKQVDGKY 161		
Qy 170 YYDDNNCKVRNFETLIADGKLHFDETGA YTDTSIDTVNK--DIVTTRSNLYKKYNQVYD 227		
Dd 162 YYDDQGVNKFNFVSGDKIYYFDETGATKDYTSKDADSSSAVSQNAITFAANNRAYS 221		
Qy 228 RSAQSFEHDHYLTAEBSWPKYLKDGKTWTOSTERKDPFLMTWMPPDOETORQYNYM 287		
Dd 222 TSAENFEADVNLTAADSWPKSLKDGKTWTESGKDDPFLMAWMPDTETKRYNYM 281		
Qy 288 NAOLGINKTVDTSNQQLNIAATIQAIEAKITLLKNTDWLRQRTISAFVKQTOSAWNSD 347		
Dd 282 NLVVIGDKTYTAETSQADLTAAAEVLQARIEQKITTEQNKTWLREALISAFVKTPQWNGE 341		
Qy 348 SEKPFDDHLONGAVLDNECKLTPYANSRYILNRPTNKGKDPRT--ADNTIGGYE 405		
Dd 342 SEKPYDHLQNGALKFDPNQSDLTPDQTSNRYLLNRPTNQTSLDGRFYTNANDPLGGYE 401		


```

Db 1296 TGAQTIRGKLYFKANGQQVKGDIVKGTGDKIRYDAKSGEQVFNKTVKAADGKTYVIGN 1355
QY 1360 NGYAV-----TGARTI 1370
Db 1356 DGVAVDPSVVKGTFKDGASGARFFYNLXGQLVTGSGWYETANHWDWYIOSGKALTGEQII 1415
QY 1371 NGOHLYFRANGVOVKGFEVTDHRGRISYYDGSNGDQIRNRFV-----1412
Db 1416 NGOHLYFKEDGHQVKGQLVTGTDGKVRYYDANSQDQAFNKSVTVNGKTYFYFGNDGTAQTA 1475
QY 1413 -----RNAQGMGFYFDNNGYAVTGAARTINGOHLFY 1442
Db 1476 GNPKGTFKDGDIRFYSMEGQLVTGSGWYENAAQQOMLYV-KNGKVLTLGLQTVGSRVYF 1534
QY 1443 RANGVOVKGEFVTDYGRISYYDANSGERV 1472
Db 1535 DENGIOAKGKAVRTSDGKIRYFDENSGSMI 1564

RESULT 5
ID GTFL_STRDO STANDARD; PRT; 1597 AA.
AC P11001;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
GN Name=gtfI;
OS Streptococcus downei (Streptococcus sobrinus).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1317;
RN [1]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC STRAIN=MFE28;
RX MEDLINE=87308014; PubMed=3040686;
RA Ferretti J.J., Galpin M.L., Russell R.R.B.;
RT "Nucleotide sequence of a glucosyltransferase gene from Streptococcus
RT sobrinus MFE28";
RL J. Bacteriol. 169:4271-4278 (1987).
CC -!- FUNCTION: Production of extracellular glucans, that are thought to
CC play a key role in the development of the dental plaque because of
CC their ability to adhere to smooth surfaces and mediate the
CC aggregation of bacterial cells and food debris.
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl)(n) = D-
CC fructose + (1,6-alpha-D-glucosyl)(n+1).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes
CC water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
CC forms of glucans.
CC -!- SIMILARITY: Belongs to the glucosyl hydrolase 70 family.
CC -!- SIMILARITY: Contains 19 cell wall binding repeats.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC
CC EMBL; M17391; AAC63063.1; -; Genomic DNA.
DR InterPro; IPR002479; Cell wall bd put.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 4.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Dental caries; Glucosyltransferase; Repeat; Signal; Transferase.
FT SIGNAL 1 38 Potential.
FT CHAIN 39 1597 Glucosyltransferase-I.
FT REPEAT 1099 1323 A repeat.
FT REPEAT 1163 1213 AC repeat.
FT REPEAT 1227 1277 AC repeat.
FT REPEAT 1292 1342 AC repeat.

```

```

FT REPEAT 1352 1399 B repeat.
FT REPEAT 1406 1455 AC repeat.
FT REPEAT 1465 1512 B repeat.
FT REPEAT 1519 1568 AC repeat.
FT REPEAT 1582 1597 A repeat (incomplete).
FT REGION 39 1050 Catalytic (approximate).
FT REGION 1099 1597 1.25 A, 2 B and 5 AC repeats.
FT REGION 1099 1597 Glucan-binding (approximate).
SQ SEQUENCE 1597 AA; 177080 MW; B9E86A200868798E CRC64;

Query Match 59.1%; Score 4578.5; DB 1; Length 1597;
Best Local Similarity 55.9%; Pred. No. 9.1e-205;
Matches 892; Conservative 216; Mismatches 339; Indels 149; Gaps 18;

QY 1 MDKVRVYKLRVKRQWTVSVASAVMTLTTLTSGGLVKADSNESKSOISNDSNTSVVTAN- 59
Db 1 MEKNERFMVKRQWTVSVASATWLASALGASVASADT-----ETVSSDSNQAVLTAQ 56
QY 60 -----BESNVIETATSKQEAASQTNRHTVTTSSTSVVNPKEVSVNPYTVGETASNG 112
Db 57 TTTNQDTQTSVAATSEQSASTDAATDQASATDQASAAEQTQGTASTDTAAQTITTA 116
QY 113 EKL-----NQTTTVDTSEAAANNISKQTTEADTDVIDSNANALQILEKLPVVKXI 165
Db 117 NEAKWVPPTENENQVFTDEMLAEA-----KNVATAESNSIPSD-----LAKMSNVKQV 163
QY 166 DGKYYYDNGNGKVRTNFTLIADGKILHFDGTGAYTDTS---IDTVNKDITVTRSNLYKKY 222
Db 164 DGKYYYDQDGNVKGNFVSVGEKIYYFDETQAYKDTSKVEADKSGSDI-SKEETFFAN 222
QY 223 NQVYDRSAQSEPHVDHYLTAEWSYRPKYILKDGKTTQSTKEDFRPLMTWMPDQSTORQ 282
Db 223 NRAYSTSAENFEAIDNYLTADSWYRPKSILKDGKTTWESSKDDFRPLLMWMPDTETKN 282
QY 283 YVYNVAOLGINKTYDDTSNQLNLNIAAATQAKTEAKITTKNTDNLQTTISAFVKTS 342
Db 283 YVYNMKNVVGIDKTYTAETSQADLTAAAEVQARIEQKITTEQNTKWLREALISAFVKTP 342
QY 343 AWNDSKPEDDHLONGAVLVNDECKLTPYANSYRILNRTNQTNGKDPYV--ADNT 400
Db 343 QWNGSEKPYDDHLQNGALKDFDNQSDLTPTQSNYRLLNRTNQTNGSLDSRFTYANDP 402
QY 401 IGGYFLLANDVDNSNPVQAEQLNLWFLMNFNIYANDPDANFDSIRVDADVNDADL 460
Db 403 LGGYELLANDVDNSNPVQAEQLNLWFLMNFNIYAKADANFDSIRVDADVNDADL 462
QY 461 LOIAGDYKAAAGIKHNDKANDHLSILEASDNDPTYLHDDGDMNMNDKRLSLSPS 520
Db 463 LQISSDYLKAAAYGIDKNNKNANNHVSIVEAWSNDPTYLHDDGDMNMNDKFRLSMLWS 522
QY 521 LAKPLNORSGMNPILTNSLVNRTDQNAETAAPVSPYSFIRAHDSVQDILADIKAIEINP 580
Db 523 LAKPLDKRSGNLPLIHNSLVNREVDRETVPSYSFARAHDSVQDILIRDIKAIEINP 582
QY 581 VVGYSFTMEEEKKAFPIYNKOLLATEKKYTHYNTALSYALLLTNKSVPVRYVYGMFTDD 640
Db 583 AFGYSFTQDEIDQAFKIYNEDLKKTDKKYTHYNVPLSYLLLTNKSIPRVYVYGMFTDD 642
QY 641 GOYMAHKTINYEATLTKARIKYVSGQAMRNQGVNSEIITSVRYGKALKATDTGDR 700
Db 643 GOYMANKTVNYDAIESLLKARMKYVAGQAMQNYQIGNGEILTSVRYGKALKQSDKGDA 702
QY 701 TTRTSGVAVIEGNPNPSLKLKASDRVVVNMGAHKQAYRPLLLTTDNGIKAYHSQEA-- 758
Db 703 TTRTSGVGVVMGNQPNFSLDGK-VVALNMGAAHANQYRMLMVSTKDGQVATYATDADASK 761
QY 759 AGLVRYTNDRCGELIFTAADIKGYANPQVSGYLVGVVVPVGA--LIKNFALRLARPHOOM 815
Db 762 AGLVKRTDENGILYFLNDDLGKVANPQVSGFLQVWVPVGAADDQDIRVAASDTASTDGK- 820
QY 816 ASVHQNAALDSRVMPFEGSFNFQAFATKKEEYTNVVIKNNVKFAEWGVTDFEMAPQYVSS 875
Db 821 -SLHQDAAMDSCRVMFEGSFNFQSFATKEEYTNVVIANNVDFKFSWGITDFEMAPQYVSS 879

```



```

QY 876 TDGSLDSVIONGYAFTDRYDLIGSKPNKYGTADDLVKAIKALHSGKIKWADWPDQWY 935
DB 880 TDGQLDSVIONGYAFTDRYDLIGSKANKYGTADQLVKAIKALHAGKIKWADWPDQWY 939
QY 936 AFPEKEVVTATVDKYGTPVAGSQIKNTLYVVDGKSSGKDQQAQKGGAFLEBLQAKYDEL 995
DB 940 TFPKQEVVTVTDFKFGPIAGSQINHSLYVTDTKSSGDDYQAKYGGAFLEBLQAKYDEL 999
QY 996 FARKOISTGVPMDSVKIKQMSAKYFNGTNIILGRGAGYVLKQDATNTVFNISDNKEINFL 1055
DB 1000 FTKKQISTGAIDPSVKIKQMSAKYFNGSNIILGRGADVLSQASNKYLVNSDDK--LFL 1057
QY 1056 PKTLINQDSQVCFSDGKGYV--YSTSGVQAKNTFISEGDKWYFYDNNGYMTGQASING 1114
DB 1058 PKTLIGQVVEGIRFDGTYGVSNTSGTEKVTDSFTEAGNLYIFYPGDGYMTGQNTKIG 1117
QY 1115 VNYFELSLQDLRALINKEDGTAYAYGNDGRYENGYYQFMGYSVYRFRNNGEMSVGLTV 1174
DB 1118 SNYIFLANGAALRNTVYTTDAQONHYGNDGKRYENGYYQFGNDSWRYPKNGVMALGLTT 1177
QY 1175 IDGQVQYFDEMGYQAKGFVTTADGKIRYFDKQSGNMYRNRIENEBGKWLVLGBDGAUV 1234
DB 1178 VDGRHVQYFEDKGVQAKDKIIVTRDGKRVYFDQHGNAVTNTFVADKTGHWYLYLKGXGVAV 1237
QY 1235 TGSOTINGOHLHYFRANGVQVKGFEVTDHGRISYYDGNSGDQJRNRFVNRNAQOGWFYFDN 1294
DB 1238 TGAQTVGQOHLHYFEANGQOVKGDFVTAKDGLFYDVSDDMTWTFIEDKAGNWFYLGK 1297
QY 1295 NGYAVTGARTINGQLLYFRANGVQVKGFEVTDGRYGRISYYDGNSGDQJRNRFVNRNAQOGW 1354
DB 1298 DGAAVTGAQTIKGKLYFKANGQOVKGDIIVKADGKIRYDAQTGEQVFNKSV-SVNGKT 1356
QY 1355 FYFDNNGYA-----VT 1365
DB 1357 YFYGSDGTAQTOANPKGTQKDGSGVLRFYNLEGQVYSGSGWYETAHEHWVYVKSGLVLT 1416
QY 1366 GARTINGOHLHYFRANGVQVKGFEVTDHGRISYYDGNSGDQJRNRFV----- 1412
DB 1417 GAQTIGNQRVYFKDNGHGVKQLVLTGNDGKRLRYDANSQDQAFNKSVTNGKTYFYGSBDG 1476
QY 1413 -----RNAQOGWFYDNNGYAVTGARTIN 1436
DB 1477 TAQTOANPKGTQKDGSGVLRFYNLEGQVYSGSGWYKNAQOGWLV-KDGKVLTLGLQVIG 1535
QY 1437 GOHLHYFRANGVQVKGFEVTDGRYGRISYYDANSGERV 1472
DB 1536 NQKVYFDKNGIQAKGKAVRTSDGKRVYFDNSGSMI 1571

RESULT 6
GTF2_STRDO
ID GTF2_STRDO STANDARD; PRT; 1592 AA.
AC P27470;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Glucosyltransferase-I precursor (SC 2.4.1.5) (GTF-I) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
OS Streptococcus downei (Streptococcus sobrinus).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_taxid=1317;
RN [1]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC STRAIN=6715 / Serotype G;
RX MEDLINE=9112327; PubMed=1704006;
RA Abo H., Matsumura T., Kodama T., Ohta H., Fukui K., Kato K.,
RA Kagawa H.;
RT "Peptide sequences for sucrose splitting and glucan binding within
RT Streptococcus sobrinus glucosyltransferase (water-insoluble glucan
RT synthetase).";
RL J. Bacteriol. 173:989-996 (1991).

```

```

CC -!- FUNCTION: Production of extracellular glucans, that are thought to
CC play a key role in the development of the dental plaque because of
CC their ability to adhere to smooth surfaces and mediate the
CC aggregation of bacterial cells and food debris.
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-
CC fructose + (1,6-alpha-D-glucosyl) (n+1).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes
CC water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
CC forms of glucans.
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
CC -!- SIMILARITY: Contains 16 cell wall binding repeats.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL; D90213; BAAL4241.1; -; Genomic_DNA.
CC HSSP; P06653; IGWM.
CC InterPro; IPR002479; Cell wall bd put.
CC InterPro; IPR003318; Glyco hydro_70.
CC Pfam; PF01473; CW binding 1; 3.
CC Pfam; PF02324; Glyco hydro_70; 1.
CC Dental caries; Glycosyltransferase; Repeat; Signal; Transferase.
CC SIGNAL 1 38 Potential.
CC CHAIN 39 1592 Glucosyltransferase-I.
CC REPEAT 1093 1142 1.
CC REPEAT 1158 1207 2.
CC REPEAT 1222 1272 3.
CC REPEAT 1287 1337 4.
CC REPEAT 1402 1451 5.
CC REPEAT 1514 1563 6.
CC REPEAT 1577 1592 7 (incomplete).
CC REGION 39 1044 Catalytic (approximate).
CC REGION 1093 1592 7 X tandem repeats.
CC REGION 1093 1592 Glucan-binding (approximate).
CC REGION 1093 1592
CC SEQUENCE 1592 AA; 176168 MW; BC0A66D079351ECP CRC64;

Query Match 58.5%; Score 4531; DB 1; Length 1592;
Best Local Similarity 55.8%; Pred. No. 1.5e-202;
Matches 889; Conservative 213; Mismatches 344; Indels 146; Gaps 19;

QY 1 MDKVRVYKLRVKRWTVSVASAVMTLTTLTSGGLVKADSNESKSIQISNDSWTSVVTANE 60
DB 1 MEKNVRFKMHVKRWVTLVSASATMLASALGASVASADTDTA-----SDDSNQTVVTGQ 56
QY 61 ESNVITEATSKQEAASSQTNHTVTTSSSSTSVVNPKE-----VSNPYTVGETASNGEKL- 115
DB 57 TTN--NQATDQTSIAATSEQASASTDAATQASAAEQTGTTASTDTAAQTITTANAEAK 114
QY 116 -----NQTTTVDKTTSEAAANNISKQTTADTDVDDSNAAANLQILEKLPNKEIDGKY 169
DB 115 WPTFENENQGFDEMLAEA-----KNVATAESDSIPSD-----LAKMSNVKQVDGKY 161
QY 170 YYYDNNQKVRNFTLIADGKILHFDGTGAYTDTSDTYNK--DIVTTRSNLYKKNQVYD 227
DB 162 YYYDQDGNVKNFAVSGDKIYYFDGTGAYKDTSKVDADKSSSAVSQATIPAAANNRAYS 221
QY 228 RSAQSFEHVDHYLTAESWYRKYTLKDGKTTWQSTEKDFRPLMTWMPDQETQROYVNYM 287
DB 222 TSAKNFEADVNYLTADSWYRPSILKDGKTTWESGKDDFRPLLMAMWPDTETKRYNYM 281
QY 288 NAQLGINKTYDDTSNQLNLAATIAQIKIEAKITTLKNTDMLROTISAFVKTSQAWNSD 347
DB 282 NKVVGIDKTYTAETSOADLTAAELVQARIQKITSNNKTWLRRAISAFVKTPQWNGE 341
QY 348 SEKFPDDHLQNGAVLYDNEGKLTVPYANSNYRILNRTPTNQTKGKDPRT--ADNTIGGYE 405
DB 342 SEKPYDDHLQNGALLFDNQDTLTPDTQSNYLLNRTPTNQTKGSLDSRFTYNPNDPLGGYD 401

```

QY 406 FLLANDVNSNPVQAEQLNHLFLMNPNIYANDPDANPDSIRVDANVNDVADLLQIAG 465
DB 402 FLLANDVNSNPVQAEQLNHLFLMNPNIYANDPDANPDSIRVDANVNDVADLLQIAG 461
QY 466 DYLKAAKGTHKNDKAANDHLSILEANSNDTPVLHDDGDNMINMDNKLRLSLFLSLAKPL 525
DB 462 DYLKAAKGTHKNDKAANDHLSILEANSNDTPVLHDDGDNMINMDNKLRLSLFLSLAKPL 521
QY 526 NORSGMNPITNSLNRDNDNAETAAPVPSYFIRAHSDSEVDLIADIKAENPNVNVGS 585
DB 522 DVRSGLNPLIHLNSLVDREVETVPSPYSPARAHSDSEVDLIADIKAENPNVNVGS 581
QY 586 FTWEEBKAFBIYNKDLATEKYTHYNTALSYALLTNKSSVPRVYGGDMFTDDCQYMA 645
DB 582 FTQEEEDQAFKIYNEDLKSDKKYTHYNTALSYALLTNKSSVPRVYGGDMFTDDCQYMA 641
QY 646 HKTINYEAJETILLKARIKYVSGQAMRNOQVGNSEIITSRYVGKALKATDTCGRTRTS 705
DB 642 NKTINYDAIESLLKAMKTVAGQAMQYQINGEILTSRYVGKALKATDTCGRTRTS 701
QY 706 GVAIEGNNPSLRLKASDRVVMGAAHKNQAYRPLLLTTDNGIKAYHSDQEA--AGLYR 763
DB 702 GGVVMGQNPFSLDCK-VVALNMGAHAHQEYRALMVSTKQGVATYATDAASKAGLYK 760
QY 764 YTNDRBELIFTAADIKGYANPQVSGYLGVVPGVAA---LIXMFAURLARPHQOMASVHQ 820
DB 761 RTDENGILYFLNDLKGAVNPQVSGFLQVVPVGAADDQDQIRVAASDSTADTGK--SLHQ 818
QY 821 NAALSRVMFEGFSNFQAFATKEEYTNVIAKNDKFAENGVTDPFEMAPQVSSTDGSGF 880
DB 819 DAAMDSRVMEFSGNSFQAFATKEEYTNVIAKNDKFAENGVTDPFEMAPQVSSTDGSGF 878
QY 881 LDSVLQNGYAFDRYDLGSGKPKYGTADDLVKAIKALHSKGIKVMADWVPQMTAFPPK 940
DB 879 QDSVLQNGYAFDRYDLGSKANKYGTADDLVKAIKALHAKGLKVMADWVPQMTAFPPK 938
QY 941 EVVTRVVDKYTPVAGSQIKNTLYVVDKSGKQQAQYGAFLFELQAKYPELPARKQ 1000
DB 939 EVVTRVTDKFKPIAGSQINHSLYVTDTKSSGDDYQAKYGAFLDELKEKYPELPARKQ 998
QY 1001 ISTGVPMDSVKIKWSAKYFNGTILGRGAGVILKQATNTYFNISDNKEINFLPKTLL 1060
DB 999 METGQAIDPSVKIKWSAKYFNGSILGRGADYVLSQVSNKYFNVA--SDTLFLPSSLL 1056
QY 1061 NQDSQVGSYDGKGVYVYST-SGYQAKNTFISEGDKWYFYDNNQYVWTCQAQNGVNYVF 1119
DB 1057 GKVVESGIRYDGKGVYVYST-SGYQAKNTFISEGDKWYFYDNNQYVWTCQAQNGVNYVF 1116
QY 1120 LNSGLQLRDAILLKNEGDTYAYYNGDGRY--ENGYQFMGSVYVHRFNNGEMSVGLTVIDG 1177
DB 1117 LENGTAIRNTIYVTDAGNSHYIYANDGKRYENENGYQF-GNDWRYFKDGNMAVGLTTVDG 1175
QY 1178 QVQYFDEMGYQAKGFVTTADGKIRYFDKQSGNMVYRNRIENEBGKWLGLBGDAVNTGS 1237
DB 1176 NVQYFDPKQGVQAKKIIVTRDGKRYFYDQHGNAVNTFTADTKGHWYLYGDKGVAVTGA 1235
QY 1238 QTINGQHLVFRANGVQKGEFVTDHGRISYYDNGSDQIRNRFVRNAQGMQFYFDNGY 1297
DB 1236 QTVGQKLYFEANGQVQKGFVTSDEGKLYFYDVSQGMWMTFTFEDKAGNFWYLGKQGA 1295
QY 1298 AVTGARTINGQLLYFRANGVQKGEFVTDYRGRISYDNGSDQIRNRFVRNAQGMQFYF 1357
DB 1296 AVTGAQTIRGQKLYFKANGQVQKGDIVKGTGDKIRYDAKSQGEQVFNKTVKAADGKTYI 1355
QY 1358 DNGYAV-----TGAR 1368
DB 1356 GNDGVAVDPSVVKGTQFKDASGALRFYNLKGQVLTGSGWYETANHDWVYIQSGKALTGEQ 1415
QY 1369 TINGQHLVFRANGVQKGEFVTDHGRISYDNGSDQIRNRFV----- 1412
DB 1416 TINGQHLVFRANGVQKGEFVTDHGRISYDNGSDQIRNRFV----- 1412
QY 1413 -----RNAQGMQFYFDNGGYAVTGARTINGQHL 1440

DB 1476 TAGNPKGTQFKDGSIRFYSMEGQLVTGSGWISNAQGMQLYV-KNGKVLTLGLQTGVSQRV 1534
QY 1441 YFRANGVQKGEFVTDYRGRISYDANSGERV 1472
DB 1535 YFDENGIAQKAVRTSDGKIRYFDENSQSMI 1566

RESULT 7
GTFD_STRMU
ID_GTFD_STRMU STANDARD; PRT: 1462 AA
AC F49331; O69383; O69386; O69389; O69392; O69398;
DT 01-FEB-1996 (Rel. 33, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Glucosyltransferase-S precursor (EC 2.4.1.5) (GTF-S) (Dextranase)
DE (Sucrose 6-glucosyltransferase).
GN Name=gTFD; OrderedLocustNames=SMO.910;
OC Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]_TaxID=1309;
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=GS-5;
RA MEDLINE=91100958; PubMed=2148600;
RX Honda O., Kato C., Kuramitsu H.K.;
RT "Nucleotide sequence of the Streptococcus mutans gtfD gene encoding
the glucosyltransferase-S enzyme.";
RL J. Gen. Microbiol. 136:2099-2105(1990).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MT4239 / Serotype C, MT4245 / Serotype e, MT4251 / Serotype f,
MT4467 / Serotype e, and MT8148 / Serotype c;
RX MEDLINE=98231843; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
RN [3]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=UAI59 / ATCC 700610 / Serotype c;
RX MEDLINE=2295063; PubMed=12397186; DOI=10.1073/pnas.172501299;
RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
Carson M.B., Prineaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,
Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental
pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
CC -!- FUNCTION: Production of extracellular glucans, that are thought to
play a key role in the development of the dental plaque because of
their ability to adhere to smooth surfaces and mediate the
aggregation of bacterial cells and food debris.
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl)(n) = D-
fructose + (1,6-alpha-D-glucosyl)(n+1).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes
water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
forms of glucans.
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
CC -!- SIMILARITY: Contains 6 cell wall binding repeats.

CC This Swiss-Prot entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use as long as its content is in no way modified and this statement is not
removed.

DR EMBL; M29296; AAA26895.1; -; Genomic DNA.
DR EMBL; D88653; BAA26103.1; -; Genomic DNA.

QY 862 GVTDFEMAPQYVSSDGSFLDSVQIONGYAFTDRYDLGISKPNKYGTADLVKAIALHSHK 921
Db 894 GVTSEMAPQYVSSDGSFLDSIIQNGYAFEDRYDLGSKPNKYGSQQDMINAVKALHKS 953
QY 922 GIKWADWPQOMYAPPEKEVVTATRVKYGTPVAGSQIKNTLYVVDGKSSGKDOQAKYG 981
Db 954 GLQVIADWPQOYNLPGEKVVATRVNDYGEYRKDSEIKNTLYIAANTKSNKGRQYQAKYG 1013
QY 982 GAFLELOQAPYELFARKQISTGVPMDPSVKIKOMSAKIFNGTNIILGRGAGVYVLQDQATN 1041
Db 1014 GAFLELAQYPSIFNRTOISNGKIDPSEKITAWKAKYFNGTNIILGRGVYVLKDNASD 1073
QY 1042 TYFNISDNKEINFLPKTLNQDSQVGSFSDGKGYVYVTSYGQAKWTFISEG-DKWYYPD 1100
Db 1074 KYFELKGNQ--TYLPKQMTNKEASTGFVNDGNGMTFYSTGYQAKNSFVQDAKGNWYYPD 1131
QY 1101 NGYVMVTGAQSTNGVNYFSLNGLQRLDAILKNEGDTYAYGNDGRYENGYYQF-MSGV 1159
Db 1132 NNGHWYGLQHLNGEVOYFSLNGVQLRESFLENADGSKNYFGLHGNRYSGYISFDNDSK 1191
QY 1160 WRHFN-NGEMSVGLTVIDGQVOYFQEMGYQAKGKFTTADGKIRYFDPKQSGNMYNRFTIE 1218
Db 1192 WRYFDASGVMAVGLKTINGTQYFDQGYQVKGAMITGSDGKRYFDDGSGMAYNRFPAN 1251
QY 1219 NEEGKWLILGEDGAAVTGSQTINGHLYFRANGVQVKGFEVTDHHRISYIDGNSGDQIR 1278
Db 1252 DKNGWYLLNSDIALVQVTINGKTYFQDQKQIKGKIITD-NGKLKYFLANGSGLAR 1310
QY 1279 NREVRNAQOWFYFDNNGYAVTARTINGQLLYFRANGVQVKGFEVTDYGRISYIDGNS 1338
Db 1311 NIFATDSQNNWYFSGDGVAVTGSQTIAKKLYFASDGQVKGSGFVTYN-GKVHYHHADS 1369
QY 1339 GQOIRNRFRVNAQOWFYFDNNGYAVTARTINGQLLYFRANGVQVKGFEVTDHHRISY 1398
Db 1370 GELQVNRFEADKGNWYLLDSNGEALTGSQRINGQVFTFRGSKQVKGVDVDERGLLRY 1429
QY 1399 YDNGSGDQIRNRFRVNAQOWFYFDNNGYA 1428
Db 1430 YDKNSGNWYKVVTLANGRRIGIDRWGIA 1459

RESULT 8

Q54178 STRGN PRELIMINARY; PRT; 1577 AA.
AC Q54178; Q54247;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Glucosyltransferase.
GN Name=gtfG;
OS Streptococcus gordonii.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1302;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Challis;
RX MEDLINE=96157084; PubMed=8586195;
RA Vickerman M.M., Sulavik M.C., Clewell D.B.;
RT "Molecular analysis of Streptococcus gordonii glucosyltransferase
phase variants.";
RL Dev. Biol. Stand. 85:309-314 (1995).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Challis;
RA Minick P., Vickerman M.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBSJ databases.
DR EMBL; U12643; AAC43483.1; -; Genomic_DNA.
DR PIR; B41898; B41898.
DR HSP; P06653; 1HCK.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.

DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 5.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1577 AA; 177805 MW; 5AE0328DC5808D18 CRC64;
Query Match 47.4%; Score 3669.5; DB 2; Length 1577;
Best Local Similarity 48.5%; Pred. No. 2.2e-162;
Matches 759; Conservative 241; Mismatches 459; Indels 105; Gaps 24;
QY 1 MDKVRVYKLRKVKKVVTVSVASAVMTLT--LSGGLVKADS-----NESKQISNDN 52
Db 2 MEKKVYKXKVKKXWVAIAVTSLLALVAPKALGESGVYADANQVTVNVKEQSAQSK 61
QY 53 TSVVTANESNIVTEATSKQEAASQTNHTVTTSST-----SVNPKEV----- 98
Db 62 DSEQTTSKATDSSQLEVKQASSKEYQOASAAINPTANEQTQODKEVESTRDSRHE 121
QY 99 ----VSNPYVTGEBTASNGEKLQNTTVDKTSSEAAANISKQTTTADTVIDDSNAANI 154
Db 122 LTQKTSDDSEKSGSQEPKVADQAESTDKTQAAALQ---AKQDSRANDQEBETTENVAKAT 178
QY 155 ILEK-----LPNVKIDGKYVY 172
Db 179 VSDKIATPKPERLPEPAQRKESITEKMLAAQAAPVNTDEHDDVLAHIKTIDGKYYV 238
QY 173 DNNGKVRTNFTLIADGKILHPD-ETGAVTDTSDITVNDIVTTRSNLYKKYNQVYDRSAQ 231
Db 239 QDGVTKQNFVAVELNGKILYFPAETGALVDSN-EVQFOQTSSLNNEFTQKNAFYTTDK 297
QY 232 SFPHVHYLTAEWYRPKYLKDGKWTOSTEKDFPLMLTWPPDQETQRYVYVYVNAQ- 290
Db 298 DIETVDGYLTADSWYRPKFLKDGKWTASTETDLRPLLMAMWPKRTQINLYNMYNQEN 357
QY 291 LGINKYTDTSNQLQNTAAATQIAKIEAKITTKNTDNLRTQISAFVKTQSAWN----S 346
Db 358 LGIG-AFESKTEQVLLTNAVQVQVKIERISKEGDTKWLRTILMSAFVKTQPNMNIKTES 416
QY 347 DSEKPFDDHLONGAVLYDNEGKLTYPVANSYRILNARTPTNQTGKDPRTVADNTTGGYEF 406
Db 417 ETGTGKDHLOGGALLYTNSDK-TSHANSRYRLNRTPTSQG--TPKTFIDKSGNGYEF 473
QY 407 LLANDVDNSNPVQAEQLNWLHFLMNFNIVANDPDANFDSIRVDAVDNVDLQLIAGD 466
Db 474 LLANDFDNSNPAVQAEQLNWLHYMNFSGIIVANDPTANFDGVRVDAVDNVDNADLLQIASD 533
QY 467 YLKAAGKTHKNDKAANDHLSILEAWSDNDTPYLDHDDGNMNMNKLRLSLFLAKPLN 526
Db 534 YFKSRYKVGESEEEALKHLSILEAWSDNDPDYNDKDTKGAQLAIDNKLRLSLFLSFMRLKS 593
QY 527 QRSGMNPLITNSLVNRTDNDNAETAAPVSPYFIRAHDSVODLIADI IKAELNPVNVGYSF 586
Db 594 IRSGVEPTITNSLDRSTENKNGERTANYIFVRAHDSVQTVIADI IREININPTDGLTF 653
QY 587 TWEIHKKAFEYNDKDLATEKKYTHYNTALSYALLLTNKSSVPRVYVYVYVYVYVYVYVYV 646
Db 654 TMDELKQAFKIYNEEDMRKADKKYQFNPTAHALMSLNKDSITRYVYVYVYVYVYVYVYVYV 713
QY 647 KTINYEATETLLKARIKTVSGQAMRNQOVG-----NSEIITSVRYGKALKATDT 697
Db 714 KSPYHDAIDALLRARIKTVAGQDMKVYMGVPREADKWSYNGILTSVRYGTGANEATDE 773
QY 698 GDRTRTSGVAVIEGNPNSRLKASDRVVVNVNGAAHKNQVPEPLLTITDNGIKAHVSDOE 757
Db 774 GTAETRTQGMVIAASNPNLKLNDKLVQNVNGAAHKNQYRVPVLLTTKDGLSRVLTDBE 833
QY 758 A-AGLVRYTNRGELIFTAADIKGYANPQVSGYGVVYVPGVGAALIKMFALRLRPHQOMA 816
Db 834 VPQSLWKKTDANGILTFDMDNDIAGVSNVQVGYLAVWVPVGAKENQDARVATSKKXNASG 893
QY 817 SVHQNA-ALDSRVMEFEGSNFQAFATKKEEYTNVVIKRVKDFBAGVGTDFEMAFQYVSS 875
Db 894 QVYESSPALDSQLIYEGFSNFQDPATRDDQYTNKVIKNNVLFKEWGVTSFELPPQYVSS 953

QY 876 TDGSLDSVIONGYAFTDRYDLGISKPNKYGTADDLVKAIALKHSKGIKVMADWVPDQWY 935
Db 954 QDGTFLDSIIIONGYAFEDRYDAMSKNKKYGLSDDLALRALHSVNIQAIADWVPDQIY 1013
QY 936 AFPEKEVVTATRVKYGTPVAGSQIKNTLYVVDGKSGKQOQAKYGGAFLELOAKYPYL 995
Db 1014 NLPKEVVTATRVNNGYTYREGAEIKEKLYVANSKTNGTDYCKYGGAFLELKAYPEI 1073
QY 996 FARKQISTGVPMPSVKIKQMSAKYPNGTINILGRGAGYVLKQDAQNTYFNISDNKEINFL 1055
Db 1074 FERVQISNGQKMTDEKIKTSKSAKYFNGTINILGRGAYVLKQDWSKEY--LSNKNGETAL 1131
QY 1056 KYTLNQDSQVGSFSDGKGYVYYSQAKNTFI--SEGDKWYIFDNGNNGYVWVTAQASING 1114
Db 1132 PKQLYNKEASTGCFVXDITNGPKFYSTSGNQAKDTFIQDENGWNYYPDNQGLYVTGAREIDG 1191
QY 1115 VNYFSLNGLQRLDAILKNEIDGTAYAYGNDGRRYENGYYQFMGSGVWRHFN--NGEMSGLT 1173
Db 1192 KQLYFMKGVQURDALQEDENGQYIYDKTGAKVLNRYYTSDQNRWYFDAGKVMARGLV 1251
QY 1174 VI--DQOVQYFDMGQYQAKGKFTTTADGKIRYFDPKQSGNMNRNFTIENEE--GKWLVLGBDG 1231
Db 1252 KIGDCQ--QYFDQNGYQVKGKVRKADGKLYFDPKDSGNVAVINRPAQGNPSPDWYVFGADG 1310
QY 1232 AAVTGSQITNGHLYFRANGVQVKGFEVTDHGRISYYDGNSGDQIRNRFVRNAGQWIFY 1291
Db 1311 VALTGHQIKQOQTLFYQDQKQVQGVVMLADKSIYFDSNGEMAVNKFAGAKNEWYY 1370
QY 1292 FDNNGYAVTGARTINGQLLYFRANGVQVKGFEVTDRIYGRISYDNGSGDQIRNRFVRNAQ 1351
Db 1371 FDQKGAVTGLKTINNQLYFDPQDQKQVQGVVTLADKSIYFDSNGEMAVNKFAGAK 1430
QY 1352 GOMFYFDNNGYAVTGARTINGHLYFRANGVQVKGFEVTDHGRISYYDGNSGDQIRNRF 1411
Db 1431 NEWYFDQDQKAVTGLQTKINKQLYFQDQKQVQGVVTLADKSIYFDSNGEMAVNKF 1490
QY 1412 VRNAGQWIFYFDNNGYAVTGARTINGHLYFRANGVQVKGFEVTDRIYGRISYDNGSG 1469
Db 1491 AEGAKNEWYFDQDQKAVTGLKTINNQLYFQDQKQVQGVVTLADKSIYFDSNGEMAVNKF 1547
QY 1470 ERVR 1473
Db 1548 DMVR 1551

RESULT 9
Q69A94 LEUME
ID Q69A94 LEUME PRELIMINARY; PRT; 1454 AA.
AC Q69A94;
DT 25-OCT-2004 (TReMBLrel. 28, Created)
DT 25-OCT-2004 (TReMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TReMBLrel. 28, Last annotation update)
DE Dextranucrase (EC 2.4.1.5).
GN Name=drrp;
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
OX NCBI TaxID=1245;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=IBT-PQ;
RA Fernandez-Pazquez J.L., Lopez-Munguia A., Olivera C.;
RT "Molecular characterization of a dextranucrase gene from Leuconostoc
mesenteroides IBT-PQ isolated from pulque."
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
RE EMBL: AY504865; AAST79426.1; -: Genomic DNA.
DR GO: 0047849; F:dextranucrase activity; IEA.
DR GO: 0046757; P:transferase activity, transferring glycosyl. . .; IEA.
DR GO: 0009250; P:glucan biosynthesis; IEA.
DR InterPro: IPR002479; CW binding.
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; CW binding_1; 1.
DR Pfam: PF02324; Glyco_hydro_70; 1.

KW Glycosyltransferase; Transferase.
SQ SEQUENCE 1454 AA; 160569 MW; 2DC7342963B6PD87 CRC64;
Query Match 47.2%; Score 3650; DB 2; Length 1454;
Best Local Similarity 52.0%; Pred. NO. 1.6e-161;
Matches 755; Conservative 210; Mismatches 398; Indels 90; Gaps 31;
QY 17 VTVSVASAVMTLTLSGLVKADSNESKSOISNDSNTSVVTANBESNVITEATSQKEA-- 74
Db 32 VTSFIQQAQADVQKNGVVVTTAVNQSNDSATTTDKS--ITTNDKATTTADTSTNDKATT 89
QY 75 -ASSQTNHTVTSSSTSVNPKKVVSNPYTVGETASNGEKLQNTTTVD-----KT 125
Db 90 TADTSTNDKATTTADTST--NDKAT-----TTADISTNNK-----ATTADTSTNNKAAT 138
QY 126 SEAAANNISKOTTEAD-----TDVIDDSN-----AANLQILEKLPNVKEIDGKYYVD 173
Db 139 ADTSNNNSATTSKDVSSSLAKSQQAIDNNSKTTDTAASFEASSK--NLKTIIDGKYYVD 196
QY 174 NNGKVRNFTLIADGKILHFD--ETGAYTDTSIDTVNKDIVTTRSNLYKYNQVYDRSAQS 232
Db 197 DNGQIKGNFATVIDGKVLDFDKTGALADTN--DYQFLEGLTSENNNYTEHNASVGTSSAS 255
QY 233 FEHVDHYLTABSWRYPKYLKDGKTWTOSTEKDPRPLMTWMPDQETQRYVYVYNAQLG 292
Db 256 YTNVDGYLTADSWRYPKDI FVNGQNWESSKDDDLRPLMTWMPDKAT---HVNYLNAMY 312
QY 293 INKT-----YDDTNSQLQNLIAAATIQAKIEAKITTLKNTDMLROTISAFVKTSQAWSND 347
Db 313 LDATETETVYTSDDSDALNKAQNIQVKEIKESIQEVOTQWLKDDISKFVDSQSNWNIA 372
QY 348 SEKFPDDLQNGAVLYDNEGKLTYPANSNYRIKARTPTNQTGKDPRTTADNTIGGYEFL 407
Db 373 SESKGTDLHOGALLYVNSDK--TPDANSDYKLLARTPTNQTG--TPLYTPTPTQGGYDPL 429
QY 408 LANVDNSNPVVQAEQLNHLHFLMNFNIYANDPDANFDSIRVDADVNDVADLLOIAGDY 467
Db 430 LANVDNSNPVVQAEQLNMGMYLLNFGSITNNDADANFDSIRVDADVNDVADLLOIADY 489
QY 468 LKAARGHKNDKAAANDHLSILEAWSNDTPLYLHDDGDNMNMNDKRLSLFLSLAKPL-- 525
Db 490 FKAAAGVVDKSDAISNQHVSIILEDVSDNDAEYVKGNDQNLSDNKLRLSLKSLTTPPVD 549
QY 526 ---NORSGMPLITNSLVNRTDDNAETHAATVPSYSPFIRAHDSQVLDLIADIKAENPNVV 582
Db 550 HYGKRSGLPEFLTNSLVNRTNDSRDNTAQPNYSFVRAHDSQVTVIAEIIKQRIIDPSD 609
QY 583 GVSFTMBEIKKAFRIYKNDLLATEKCYTHYNTALSYALLTNKSSVPRVYVYGDMTDDQ 642
Db 610 GLSPTMDQLTEAFKIYNADQLKTDEFTQYNI PSTYATILTNKDTVPRVYVYGDMTDDQ 669
QY 643 YMAHKTINYEAIETLLKARIKYVSGGQAMRNQOV--GNSE-----TITSVRYGKGALK 693
Db 670 YMATKSLYDAIDTLKLSRIKYVSGGQTMKMKYMGQDSSMAADSRYGILTSVRYNGAWT 729
QY 694 ATDTGDRTRTSQVAVTEGNNPSRLKASDRVVNMGAAHKNQAYRPLLLITDNGIKAYH 753
Db 730 ATDAGTNETRTHGIAVIESNNPDLKLSSTDQVVDVMDGIAHKNQAYRYPALLTTKGDIDTV 789
QY 754 SDQEA--GLVRYTNDRGELIFTAADIKGVANPOVSGYLGVMVVPVGAALIKMPALRLARH 812
Db 790 SDSVDSQSLIRYTNNGQLIFNSSDITVGTANPQASGLAVVPVQAS-----DTQARTE 844
QY 813 QQMA-----SVHQNAALDSRMFEGFSNFQAFATKKEEYTNVVIKAVNDKFAEAGVTDF 866
Db 845 SSTATTTDQGLHNSAALDSQVIESFNSFQSTPTTEAEYANVQIANNTDLYKSWGITNF 904
QY 867 EMAYQVSSSTGCSFLDSVIQNGYAFTRDYDLGISKPNKYGTADDLVKAIALKHSKGIKVM 926
Db 905 EFPQYRSSTDSFLDSIIQNGYAFTRDYDLGIFNTPTTKYGTVDQLRTAKALHATGIIKAM 964
QY 927 ADWVPDQWYAFPEKEVVTATRVKYGTPVAGSQIKNTLYVVDGKSGKQOQAKYGGAFLE 986

Db 965 ADWVPDQIYNLTGKEVAVQVRVNNSGIYNQDSVINTKLTVASQTVGGG-EYQALYGGFELD 1023
QY 987 ELQAKYPELFARKQISTGVPMDFSVKIKQWSAKYFNGTINILGRGAGYVLKQDQATNTYFNI 1046
Db 1024 EIKKLYPSLFERKQISTGVPMDEASEKIKWSAKYFNGTINILGRGAGYVLKQDQATNTYFNI 1083
QY 1047 SONKEIN-FLPKTLNQDSQVGF-SYDGGYVYISTSGYQAKNFTISEGDK--WYFYDNN 1102
Db 1084 STSSNSSVFLPKQLTNEESNTGFISTDG-GMTYISTSGYQAKDFTIQD-DKSNWYFDKN 1141
QY 1103 GYMVTGAQINGVNYFLSNGQLRDAILKNEDGTVAAYGNDGRRYENGYYQFMSGVWEH 1162
Db 1142 GYMTYGFQVNDNNYFLPENGIELQDAILLEDKGDVYFNQYKQITIDGYMLANKTWRY 1201
QY 1163 FN-NGEM-SVGTJ--VIDQ--VOYDEMGYQAKGFVTTADGKIRYFDKQSGNMVNRNF 1216
Db 1202 FDKNGVMAAGLTTVTVDQKHQIYFDKNGIQVKGTSVKADGKLYRFTDSEGMVNRNF 1261
QY 1217 IENESEKWLILGEDGNAVTGSGQINGOHLYFRANGVOVKGEFVTDHGHRIISYDNGSGDQ 1276
Db 1262 GEKTDGTSYFGADGIATVTAQITISQKLFDDADGQIQKGEATDKKGHGHYDADSGEM 1321
QY 1277 IERNFRVNAQGWFFDNNGYAVTGARTINGOLLYFRANGVOVKGEFVTDHGHRIISYDNG 1336
Db 1322 TTNREKLSDGSMVFNKGNVVTGAQVINGQHLFFESNGNQVKGREYATDQKMYIDA 1381
QY 1337 NSGDIQIRNRFRVNAQGWFFDNNGYAVTGARTINGOHLYFRANGVOVKGEFVTDHGHRI 1396
Db 1382 DSGDMVTNRFRISDGMAYFGANGVAVTGAQNINGQQLYFDTKGHVKGAAVAQADSGQ 1441
QY 1397 SYDNGSGDIQIRN 1409
Db 1442 KYDANSGBELIKS 1454

RESULT 10
Q9LCH3_STOR PRELIMINARY; PRT; 1575 AA.
AC Q9LCH3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Glucosyltransferase.
GN Name=gtfR;
OS Streptococcus oralis.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=1303;
RW [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC10557;
RX MEDLINE=20231779; PubMed=10768934;
RX DOI=10.1128/TAI.68.5.2475-2483.2000;
RA Fujiwara T., Hoshino T., Ooshima T., Sobue S., Hamada S.;
RT "Purification, characterization, and molecular analysis of the gene
RL encoding glucosyltransferase from Streptococcus oralis.";
RL Infect. Immun. 68:2475-2483(2000).
DR EMBL; AB025228; BAA95201.1; -; Genomic_DNA.
DR HSSP; P06853; 1HCX.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; 1CW binding_1; 4.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1575 AA; 176792 MW; 772A26B4D7C2E543 CRC64;
Query Match 46.5%; Score 3597.5; DB 2; Length 1575;
Best Local Similarity 48.0%; Pred. No. 5e-159;
Matches 749; Conservative 236; Mismatches 475; Indels 101; Gaps 24;
QY 1 MDKVRVYKLRKVRWTVSVASAVMTLTT-----LSGGLVKAD----- 39

Db 2 MEKIHVKHVKQKNVAIALATLALIVAPKVLGLSEGVVHADDDVKQVVQVEBPATQTS 61
QY 40 ---SNEKSOISNDSTSVVTANEESNVITEATSKQEAASQTNHVTSSSS--TSVNP 95
Db 62 PQOQTPAQAKIASOEAEKVTADKVTDDVAASEKPAKPAENTEATVQNAQPAKPADT 121
QY 96 KEVSNPYTVGTASNGEKLQQTITVDKTSBAAANNISKQITTEADTD-----VIDDS 148
Db 122 KEASTEKAABAEV---KAANAITEIPKTEVADQNKQARPPTAQOQEGDKREKTAVE 177
QY 149 NAANIQLLEK-----LPNVKEIDGKYYVD 173
Db 178 IVANPKVAKDKLPBPGSKQGAIAERWADQAPAPVADHDDDLVLSHKTIDGKNYYV 237
QY 174 NNGKVRTFTLTADGKILHFD--ETGAYTDTSDTVNKKDVTTRSNLYKKYQNVYDRSAQS 232
Db 238 DDTGTVKQNFVAVELNKRILYFADTETGALVDSN-EYQFQQTSSLNNEFSQKNAPYGT 296
QY 233 FEHVDHYLTAEWYRPKILKDGKTWTOSTEDFRPLMTWMPDQETQRYVNYMNAQ-L 291
Db 297 IETVDGYLTADSWYRPKILKDGKTWTASTETDLRPLLMWMPDKRTQINLYNMNQGL 356
QY 292 GINKTYDDTSNQLNIAAATIQAKIEAKITTLKNTDMLROTISAFVKTOSAWN----SD 347
Db 357 GAG-AFENKVEQALLTGASQQVQVKIEKIGKEGDTKMLRTLMGAFVKTQPNWNKTES 415
QY 348 SEKPPDDHLQNGAVLYDNEGLTPYANSNYRILNRTPTNOTCKDPRYTADNTIGYEF 407
Db 416 TTGTCKDHLQGLLYTNEK-SPHADSKFRLLNRTPTSTQ--TPKYFIDKSNNGYEF 472
QY 408 LANDVDNSPVVQABQLNWLHFLMNFNIYANDPDANFDSIRVDAVDNVDADLLQAGDY 467
Db 473 LANDFDNSPVAQABQLNWLHVMNFGSIVANDPTANFDGVVDAVDNVDADLLQIASY 532
QY 468 LKAAGIHKNDKAANDHLSILEAMSDNTPYLHDDGDNMNMNKLRLSLSLAKPLNQ 527
Db 533 FKSRVKGVESEEAIKHLSILEAMSDNDPDYKDKTKGAQLADNKLRLSLYSFMRNLGI 592
QY 528 RSGMPLITNSIVNRTDDNAETAAPVPSYFFRAHDEVDLADIKAENINNVGYST 587
Db 593 RSGVEPTITNSLDRSEKNGERMANYIFVRAHDESVQTVIADIIRINENPTDGLTFT 652
QY 588 MEEIKKAFIYNKOLLATEKTYTHYNTALSYALLITNKSSVPRVYVYVDMFTDDGQYMAH 647
Db 653 MDLQKAPKIYNEDMRKADKKTQNIPTAHMLSNKDSITRVYVYVYVYVYVYVYVYV 712
QY 648 TINYEAITLLKARIKYVSGGQAMRNQVQV-----NSGIITSVRYKGAALKATDTG 698
Db 713 SPYHDAIDALLRARIKYVAGGQDMKVTYMGVPREADKWSYNGILTSVRYGTGANEATD 772
QY 699 DRTTTSQVAVIEGNNPSLRKASDRVVVNGAAHKNQAYRPLLLTTTNGIKAYHSDQBA 758
Db 773 TAETRTQGMVAVIASNNPNLKLNEWDKLQVNGAAHKNQYRPLVLTTKDGRISRYLTDE 832
QY 759 -AGLVRYTNDRGELIFTAADIKGVANPOVSGVGVVVPVGAALIKWFAALRLARPHQMA 817
Db 833 PQSLWKTNDANGILTFDMDIAGSNVQVSVGLAVVVPVGAQADQARTTASKKQNASQ 892
QY 818 VHQ-NAALDSRVMPFGFSNFOAFATKKEEYTNVVIKNDKPAEKGVTDFEMAPQVYST 876
Db 893 VYESSAALDSQLIYEGFSNFQDFATRDQYTNKVIKVNVLKFEKGVTSFELPPQVYVSSQ 952
QY 877 DGSFLDSVIQNGYAFTRDYDLGISKPNKYGTADDLVKAIKALHSGIKVMADWDVQMYA 936
Db 953 DGTFLDSIIQNGYAFEDRYDMAMSKNNKYGSLKOLLNLRALHVSNIQAIADWVDPQYN 1012
QY 937 FPEKEVWTATRVKYGTPVAGSOIKVTVYVVDGSSGDKQQAQYGGAFLEELQAKYPELF 996
Db 1013 LFGKEVWTATRVNNGYTYREGAEIKELKYVANSKTNETDFQKYGGAFLDELAKYPELF 1072
QY 997 ARKQISTGVMPDPSPVKIKQWSAKYFNGTINILGRGAGYVLKQDQATNTYFNISDNKEIN 1056

```
Db 1073 ERVQISGQKMTTDEKITKSAKYFNGTNILGRGAYVYLKDWASNDYLT-NRNGEI-VLP 1130
Qy 1057 KTLNQNDSQVFSYDGGKYYVYSTSGYQAKNTFI--SEGDKWYFNNNGYMTGAOSINGV 1115
Db 1131 KQLVNKSYTGFVSDANGTKFYSTSGYQAKNSFIQDENGNNYYFDKRGYLVGTGAHEIDGK 1190
Qy 1116 NYFELNGLQRLDAILKNEGDGTAYVYGNDRRYENGYQFMSGVWRHEN-NGEMSGLTV 1174
Db 1191 HVFUJNGIQRDSIREDENGNQYYYDQTGAQVNLRYVTTDQGNWRYFDAGVWARGLVK 1250
Qy 1175 I-DGQVQYFDEMGYQAKGFVTTADKIRYFDKQSGNNRNFIEREE-GKWLVLGEDGA 1232
Db 1251 IGDGQ-QFDENGYQVKGKIVSAKDGKRLYFDKQSGNAVINRFAGDNPSDWYFYGVZFA 1309
Qy 1233 AVTGSQTNGHLYFRANGVQVKGFEVTDHGRISYYDGNSGDQIRNRFVNAOQWYFY 1292
Db 1310 KLTGLQKIGQOTLYFDQDGKQVKGKIVLSDKSIYFPDANGEMA VGFAGKANEWYFY 1369
Qy 1293 DNNGYAVTGARTINGQLLYFRANGVQVKGFEVTDYGRISYYDGNSGDQIRNRFVNAOQ 1352
Db 1370 DKTGKAVTGLOKIGKQOTLYFDQDGKQVKGKVVTTADKSIYFPDADSGEMAVGKFAEGAKN 1429
Qy 1353 QWYFDDNGYAVTGARTINGHLYFRANGVQVKGFEVTDHGRISYYDGNSGDQIRNRFV 1412
Db 1430 EWIYFDQTKAVTGLOKIDKQTLYFDQDGKQVKGKIVLSDKSIYFPDANGEMATNKFV 1489
Qy 1413 RNAQOQWYFDDNGYAVTGARTINGHLYFRANGVQVKGFEVTDYGRISYYDANGSERV 1472
Db 1490 EGSQNEWYFDQAGKAVTGLOQVGGQOTLYFDQDGKQVKGK-VVDVNGVSYFDANGSDWA 1548
Qy 1473 R 1473
Db 1549 R 1549

RESULT 11
Q9WXJ5_9STRE PRELIMINARY; PRT; 1512 AA.
AC Q9WXJ5;
DC 01-NOV-1999 (TremBrel. 12, Created)
DT 01-NOV-1999 (TremBrel. 12, Last sequence update)
DT 01-JUN-2003 (TremBrel. 24, Last annotation update)
DE GTP-S.
GN Name=gTfT;
OS Streptococcus criceti.
OG Plasmid pAM1.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=1333;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=HS-6;
RA Inoue M., Fukui K., Miyagi A.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB026123; BAA77237.1; -; Genomic_DNA.
DR HSP; P06654; IMPE.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW binding_1_2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Plasmid.
SQ SEQUENCE 1512 AA; 167145 MW; 4C03D9CBC601FC14 CRC64;

Query Match 45.5%; Score 3523.5; DB 2; Length 1512;
Best Local Similarity 48.0%; Pred. No. 1.3e-155;
Matches 739; Conservative 239; Mismatches 459; Indels 101; Gaps 34;

Qy 1 MDKRYKRLKRYKRWVTSVASAVMTLTLSGG---LVKADSNESKSQISNDNTS--- 54
Db 1 MERKLYKHLKVKQWVTIAVASAGLASVIGAGASQTVSADDICNGASASAEQNTSASQ 60
Qy 55 ---VVTANEESNVTIETASKQEAASSQTNHTVTTSSSSSTSVNPKVSVN-----PYTVG 106
```

```
Db 61 NKEVDSDAQAQTAQAKATSEQAAVSS-----VDTNSEIDQVQNDGVSANNQPEAPAPQ 115
Qy 107 ETASNG-----EKLQNTTTVDKTTSEAAANNISKQTTTEADTDVDDSSNAANLQILEKLPN 161
Db 1131 AAASNNNTATSEBANTNTAVSEAAPAAEN---RTAEKGADLSDOEEAAL-----SLDN 167
Qy 162 VKEIDGKYYYDNGKVRTNFTLIADGKILHFD-ETGAYTDTSDTVDNKDIVTTRSNLYK 220
Db 168 IKKYDGYKYVMADGSYKKNFAITVKGQLLYPDAETGALSSTSTVSFSGQLTPLVSD-FS 226
Qy 221 KYNOYVDRSAOSFEHVDHLYTAESWYRPKYILKQKWTQSTQTEKDFRPLLLMTWPDQETO 280
Db 227 INNKAFOSSAKSFEVLDGYLTAESWYRPTKLENGKTWDSKETDLRPLVLSWNPDKDTQ 286
Qy 281 RQYVNYNMAQINGINKTYDDTSNQLNIAAATIQAIEBAKITTLLKNTDMLRQTISAFVKT 340
Db 287 VAYLNYMSKALGGKEEFTTKTSQTALNTAAEMIQWKIEQRISKEQGTAWLRDAMAFAVAT 346
Qy 341 QSAWNSSEKPPD--DHLONGAVLYDNBEGKLTPTANSNYRILNRPNTQTKGKDPRTYAD 398
Db 347 QSRWNSDSEQ-FDKNDHLQGGALLYTN-NKLTWADSKYRLNRPTRQDGK--PHYSKA 402
Qy 399 NTIGSYFELLANDVNSNPVQASOLNWLHFLMNFNIYANDPDANFDSIRVDAVDNDA 458
Db 403 DEYGYFELLANDVNSNPVQAEMLQIHYLMNWSIWMGDKNANFGIRVDAVDNDA 462
Qy 459 DLLQIAGDYLKAAKGIHKNKKAANDHLSILEASDNDPTFYLHDDGDNNINMKNLRLSLL 518
Db 463 DTQLQYTNFYFNAVYGVDSKSEAQALAHISILEAWSYNDNYNQDNTNGAALMDNGLRLSLL 522
Qy 519 FSLAKPLNQRS-GWNPPLITN--SLVNRDADNAETAAPVPSYFIRAHDSVQDLADIKA 575
Db 523 YTLRPLSERTPGLSTLTKSEYGLTDRTKDQYGTQPSYFVFRAHDSVQTVIAQIIKE 582
Qy 576 BINPVVGYSPTEBEEKKAFIYNKDLLATKKYTHYNTALSYALLLNKSSVPVYGD 635
Db 583 KIDPTDGTFTLDQLQKQAFDIYNKDMNSVBKHYTHYNIIPAAYAVLMSNMESVTRVYGD 642
Qy 636 MFTDDGQYMAHKTINYEAETLLKARIKYVSGGQAMRNQO-----VGN 678
Db 643 LFTDDGQYMETKSPYDAINTLLRARIRYAGGQTMHKA YTPSAAMKAKNPDSGVLGN 702
Qy 679 SEITSVRYKGALKATD-TGDRTRTSGVAVIEGNPSLRKASDRVVVVMGAHKHQA 737
Db 703 SEVLVSVRFQGVMSADDMTGGLAKTSGMFLANNPELELDANEELKVNAGKIHAGQA 762
Qy 738 YRPLLLTTDNGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKGYANPOVSGYLVWVPVG 797
Db 763 YRPLLLTTDGLQKLYLNSD--TNLTKVADKQGFITFKGSEIKGYKQVEVNGVLSVWVPVG 821
Qy 798 AALIKMFALRLARPHQQMASVHONA-----ALDSRVMEFGFSNFQAFATKBEYT 847
Db 822 AKSDQ--DIRVA-----ASTKANAKGSKSVTASQALDSQLIYEGFSNFQFVQKDAQYT 873
Qy 848 NVVIAKNVDKFAEWGVTDFEMAPQVSVSTGSDSLDSVLQNGVAFTRDRLDGLSKNKYGT 907
Db 874 NKKIAENTDLFKAWGVTSFENAPQVVSATDGTDFLDSIIQNGVAFSDRYDLAMSKNKYGS 933
Qy 908 ADDLVKAIKALHSGIKVMADWVDPQYAFPEKEVVTATRVDKYCTPVAGS01KNTLYVY 967
Db 934 KEDLANALKALHAAGIQAIADWVPDQIYQLPGKEVVVTASRVNDRVGRVKIDQPMVWKLIA 993
Qy 968 DGKSGKQQAQYGGAFLEELQAKYPELFAFKQISTGVPMDFPSVKIKQWSAKYFNGTNIL 1027
Db 994 NTKSGKDFQAKYGGEFALQKQYPEMFTAKMISTGKPIDSSVLKWSAQYFNGTNVL 1053
Qy 1028 GRGAGYVLKQATNTYFNISDNKEINFLPKTILL-NQDSQVGSFYDGKGVYVYSTSGYQAK 1086
Db 1054 GRGTDYVLSDSGTGKYFTVNEKGE--FLPAVLTDGKEAKTGYNDGKGMGYFTTASQAK 1111
Qy 1087 NTFISEGDKWYFNNNGYMTGAOSINGVNYFFLSNGLQRLDAILKNEGDGTVAAYGNDGR 1146
```


Db 1112 SDFVTVAGNTYFDYTGHMVTVGNGINTKPYFVFLPNGVMLKDVAWMBDDRGSRVYVYKGTGV 1171
 QY 1147 RYE----NGYOFMSGV----WRHFN--GEMSVGLTVIDGOVQVYFDEMGOAKGKEVTTA 1197
 Db 1172 MYKSGRNNEFWAMTDSKGQLRFRHFDNYGFMVGLVTHGNVQYDEEFGVQVKGDFVTDK 1231
 QY 1198 DKIRYFDKQSGNMVNRRIENEEGKWLVLGBDGAATVGSQTINGOHLFYRA-NGVQVKG 1256
 Db 1232 AGCTRYFDKNTGNLKGQFF-NONGHWYSDOGLIAKGAQTIKGOKLYFDKTAQVKG 1290
 QY 1257 EFTVHHGRISYDGNSGDQIRNRFVRNAQGWFFDNGGYAVTGARTINGQLLYF-RAN 1315
 Db 1291 DFTVDKGNFTFYSGTDGLAVSTFTSGTNNAFYADENGVHAKGKTINGOKLYPDTKT 1350
 QY 1316 GVQVKGFEFTDRYGRISYDGNSGDQIRNRFVRNAQGWFFDNGGYAVTGARTINGO 1373
 Db 1351 GQOAKGRFVRDAKG-LRFYDAUTGALVTNSFLETRAGSNQWYMGADGVAVRGHQTIOQR 1409
 QY 1374 HLYFRA-NGVQVKGFEFTDRYGRISYDGNSGDQIRNRFVRNAQGWFFDNGGYAVTGA 1432
 Db 1410 HMYFDAETGQAKGIWVTDANGKVFYDANTGDRVNQFVL-VNGSWYFFGYDGAATGF 1468
 QY 1433 RTINGOHLFYFRANGVQVKGFEFTDRYGRISYDANSGE 1470
 Db 1469 RDIRGOHLFYFNPDTGQAKGTTVKID-NRIYTFDADSGE 1505

RESULT 12
 Q56CX8 95TRE
 ID Q56CX8_95TRE PRELIMINARY; PRT; 1506 AA.
 AC Q56CX8;
 DT 10-MAY-2005 (TremBLrel. 30, Created)
 DT 10-MAY-2005 (TremBLrel. 30, Last sequence update)
 DT 10-MAY-2005 (TremBLrel. 30, Last annotation update)
 DE Glucosyltransferase-T.
 GN Name=gtft;
 OS Streptococcus sobrinus.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1310;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=B13N;
 RA Kuwahara N.;
 RL Submitted (MAR2005) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY966490; AAX76986.1; -; Genomic_DNA.
 KW Transferase.
 SQ SEQUENCE 1506 AA; 167016 MW; 94B882EF2C17C451 CRC64;

Query Match 45.4%; Score 3512.5; DB 2; Length 1506;
 Best Local Similarity 46.5%; Pred. No. 4.3e-155;
 Matches 716; Conservative 262; Mismatches 461; Indels 101; Gaps 31;

QY 1 MDKKVRYKLRKVRWTVSVASAVMTLTTLSGGL-----VKADSN 42
 Db 1 MERKLHYKLHKVKKQWTVIAVASGLASTVAGAGLSQTVSADDLAKEQAQASQQAANAQ 60
 QY 43 SKSQISND-SNTSVVTANESNVITEATSKQEAASSQTNHTVTTSSSTSVVNPKEVSN 101
 Db 61 NEDEVASDAADTASAKATSEKEVWQSDTNSQVETKQDQAKESADAVAK-QAPQAG 119
 QY 102 PYTVGETASNGEKLQNTTVDKTEAANNISKQTTADTDVIDDSNAANLQILEKLPN 161
 Db 120 PATTSQVASSSSSVAPSEADK---AAGSVSQNEEAAAL-----SLAN 161
 QY 162 VKEIDGKYTYDNNNGKVRNTFTLIADGKILHFD-ETGAYTDTSIDTVNKRDIVTTRSNLYK 220
 Db 162 IKKIDGKYTYVWADGSYKKNFAITVDGQWLYFDKTAAGLSSTSTVSFSGGLTPIVSD-PS 220
 QY 221 KYNQVDRSAGSFEHVDHYLTAESWYRPKYILKQGTWTSQTEKDFRPLMTWTPDQETO 280
 Db 221 VNNKAFDSSEKSFELVDGYLTAESWYRPAKILENGKTVWDSKETDLRPLMSWNPNDQTO 280

QY 281 ROYVNMNAQLGINKTYDDTSNQLQINIAAATQAKIEAKITTLKNTDMLROTISAFVK 340
 Db 281 VAYLVNMSKALGKKEEFTTETSQTLNTAAELIQAKIEARVSKEQCTKWLREAAAFVAT 340
 QY 341 QSAWNSDSEKPPD--DHLQNGAVLYDNEGLKLPYANSNRIILNRTPTNOTGKKDPYTAD 398
 Db 341 QSRMKNKDEQ-YDKADHLQGGALLYTN--NNLTWANSNWRLLNRTPTRODGK--THYSKA 396
 QY 399 NTIGGVFEFLANDVNSNPVQAEOLNWLHFLMNFIGNIVANDPDANFDSIRVDADVNDVA 458
 Db 397 DKYGGYFEFLANDVNSNPVQAEMLNQIHYLMNWEIVMGDKNANFDGIRVDADVNDVA 456
 QY 459 DLLQIAGDYLKAAKGHIHKNDKAANDHLSILEAWSNDTPLYLHDGDNMINMKNKLSLL 518
 Db 457 DTLQLYTNVNSVYGVNKGSEQAALAHISVLEAWSYNDNDYNDQDTNGCAALAMDNGLRFSLL 516
 QY 519 FSLAKPLNORS--GMPLITN--SLVNRDNDNAETAAPSYSFIRAHDSVQDLADIKA 575
 Db 517 YLTRPINERTPGMSTLIKSEYGLTDRTKNDKYGDTQPSYFVRAHDSVQTVIAQIIKE 576
 QY 576 EINPNVGVYSFTMBEIKKAFELYNKOLLATEKKYTHYNTALSYALLTNKSSVPRVYGD 635
 Db 577 KIDPTDGTFTFLDQLQKAFELYNKDMNSVNKHYTHYNIPAAAYAVMLSNMESVTRVYGD 636
 QY 636 MFTDDGQYMAHKTINYEATILLKARIKYVSGQAMRNOQ-----VGN 678
 Db 637 LFTDDGQYMAKSPYDAINTLLRARIYAAGQIMEHNSYKPSAAMKAADHAGNVGN 696
 QY 679 SEIITSVRYGKALKATD-TGDRTRTTRSGVAVIEGNPSLRKASDRVVVNVNNGAAHNOA 737
 Db 697 SEVLVSFVGQDVMGADMTGKLAKTSGMFTLISNNPELELDVNEEIKVNVGKIHAQOA 756
 QY 738 VRPLLLTTDNGIKAVHSDQEAAGLVRYNYDRCELFTAADIKGYANPQVSGYLVGVVPG 797
 Db 757 VRPLLLTTDKGLQKYLNDSTK-LTKIADKQGITFKGSEIKGYQVENVGLSVLVVPG 815
 QY 798 AALIK--MFALRLARPHQOMASVHQAALDSRMFEGFSNFQAFATKBEYTNVVIAXNV 855
 Db 816 AKADQDIRVAPSTAAGEKAKTYTASQALESQLIYEGFSNFQDFVQKDSQYTNKKIAENT 875
 QY 856 DKFAEWGVTDFEMAPQYVSTDSGLDSVIONGYAFTDRYDILGIGSKPNKYGTADDLVKAI 915
 Db 876 DLFKAWGVTSPEMAPQYVSATDGTGLDSIENGAYFTDRYDLAMSKNNKYGSKEDLANAL 935
 QY 916 KALHSKGIKWADWVDPQOMVAPPEKEVVTATRVDKVGTFPVAGSQIKNTLYVVDGSSGD 975
 Db 936 KALHAAGIQALADWVDPQIYQLPGKEVVTASRVNDYGRVKVDQPLVEKLYLANTKSSGD 995
 QY 976 QQAKYGAFLBELOAKYPPELFARKQISTGVPMDDPSVKIKOMSAKYFNGTNILGRGAGYVL 1035
 Db 996 FQAKYGGFELAELOKYPPEMTTKMISTGKTIDPSVKLEWSAKYFNGTNVLDRTGYIL 1055
 QY 1036 KQATNTVFNISDNKEINFLPKTL--NODSQVGSYDGKGVYVYSTGSAKNTFISEGD 1094
 Db 1056 SDEGTGKYFTV--NEKGDPLASLTGNKDAKTGFYNDGKGIYVYTAGNKARSASFTEAG 1113
 QY 1095 KWYFDNNGYMTGAQSIINGVNYVFLSNGLOLRDALIKNEDGTAYAYGNDGRYENGYQ 1154
 Db 1114 NTYFDYTGHMVTPGNVINTKPYFPLPNGIMLKDAIKDEKRSYVYKGTGYMYKGG---- 1170
 QY 1155 FMSGVW-----RHFNN--GEMSVGLTVIDGOVQYFDEMGOAKGKEVTTADGKI 1201
 Db 1171 -RDNEWFAMTDSKQMRFRHFDYGFMSIGLVTINQVYDENGFOVKGEFVTDQDGT 1229
 QY 1202 RYFDKQSGMYNRFTIENEEGKWLVLGBDGAATVGSQTINGOHLFY-RANGVQVKGFEFT 1260
 Db 1230 RYFDQSGNLVKGQFL-NKDGWNYLDDQGLVAKGAQTIKGOKLYFDTKGTQVQVKGDFVT 1288
 QY 1261 DHGRIEISYDGNSGDQIRNRFVRNAQGWFFDNGGYAVTGARTINGQLLYF-RANGVOV 1319
 Db 1289 DKQGNTPFYSGDTGDLILGQFFSTGNNNAWFADENGHVAKGAKTIRGOKLYFDTTQGOA 1348
 QY 1320 KGEFVTDYGRISYDGNSGDQIRNRFVRNAQGWFFDNGGYAVTGARTINGOHLFY 1377


```
Db 1349 KGRFRDDKG-VRYADATGLVTVNAFLETKAGSNQWYMGADGAVAVKGNQTIKNQHWYF 1407
Qy 1378 RA-NGVQVKGEFVTDHRGRISYDNGSGDQIRNRRVNRNAQGWQFYDNNGYAVTGARTIN 1436
Db 1408 DAETQQQAKGIIVTDANGRKIFYDFTGSRVVNQFVL-VNGNWYFFGYDGAAVTGPHDIK 1466
Qy 1437 GOHLYFRANGVQVQGEFVTDYGRISY-YDANSGERVRIN 1475
Db 1467 GOHLYFNSDGTQAKG--TTVKIGNRSYTFDAHTGELTSVH 1504

RESULT 13
Q55265 STRSL
AC Q55265; STRSL PRELIMINARY; PRT; 1577 AA.
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)
DE Glucosyltransferase precursor.
GN Name=gtfm;
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=95122197; PubMed=7822030;
RA Simpson C.L., diffard P.M., Jacques N.A.;
RT "Streptococcus salivarius ATCC 25975 possesses at least two genes
RT coding for primer-independent glucosyltransferases.";
RL Infect. Immun. 63:609-621 (1995).
DR EMBL; L35928; AAC41413.1; -; Genomic_DNA.
DR PIR; T30858; T30858.
DR HSSP; P06653; 1H8G.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR004829; Csurface_antigen.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 4.
DR Pfam; PF02324; Glyco_hydro_70; 1.
DR ProDom; PD153432; Csurface_antigen; 1.
KW Signal; Transferase.
FT SIGNAL 1 38 Potential.
FT CHAIN 39 1577 Glucosyltransferase.
SQ SEQUENCE 1577 AA; 175290 MW; 3EPB989A7D3A7BF3 CRC64;

Query Match 45.3%; Score 3506; DB 2; Length 1577;
Best Local Similarity 45.8%; Pred. No. 9.2e-155;
Matches 733; Conservative 239; Mismatches 468; Indels 162; Gaps 28;

Qy 1 MDKVRVYKLRVKRWVTVSVASAVMTLTLSSGLV----KADSNESKS-----QISN 49
Db 1 MENKVRFLKVKKWNVIGVTT--LSWVALGSLLAGKGVKVEADTSAPNGDGLQQLSE 58

Qy 50 DSNTSVVTANESNVITEATSKQ-----EASQTNHVTVTSSSTSV 92
Db 59 DGTASLVT---TTVTVEQASQAASVAVATASVSHETSFOAATSASVQEAATAQAQTSVP 114

Qy 93 VNPKEVSNPVYVGETASNGEKLO----- 116
Db 115 ASQEVAVSS-----QTSSSGQETQTEQVSGQTSQVAGQTSQAQTSPTSVEQAARPVLT 169

Qy 117 -----NOTTVVDKTSSEAAANNI-----SKQTTREAD 141
Db 170 NAAPAAATADSTIRINANRNTNITATSGTTPNVTITGPNTPKPNVTVTSPNGTRPN 229

Qy 142 TDVIDDSNAANLQILEKLIP-----NVKEIDGKYIYDNNNGKVRT 180
Db 230 VTIVTQPNQPNKVPQSPQPNKVPQPNKVPQPNKVPQPNKVPQPNKVPQPNKVPQPNKVP 288

Qy 181 NFTLIADGKILHFDGTAYTDTSIDTVNKDITVTTRESNLNKKYQNVDRSAQSFEHVHYL 240
```

```
Db 289 NAAIELDRLYFDEFTGAMVDQSKPLRYADAI PNNS-IYAVNQAYDTSSKSEFHLNPL 347
Qy 241 TAESWYRPKYILKDGKWTQSTEXDPRPLMTWMPDQETQRQYVNYNMAQGLINKTYDDT 300
Db 348 TADSWYRPKYILKDGKWTASTEXDYRPLMTWMPDKVTQVNYLYNMQOQFGNKTYTDD 407
Qy 301 SNQQLNIAAATIQAKTEAKITTLKNTDNLQTTSAFVKTQSANNSDSEKPP---DDHLQ 357
Db 408 NMSYDLAAAEVTVQGIIEERIGREGNTTWLRLQMSDFIKTPGMWSESEDNLLVGDHLQ 467
Qy 358 NGAIVYDNEGKLTPIYANSYRILARTPTNQTGKDPRTADNTIGGYEFLLANDVNSNP 417
Db 468 GGALTFLN-NSATSHANSDFRLMRTPTNQGTGR--KYHIDRSNGYELLANDIDNSNP 524
Qy 418 VVQAEQLNLHFLMNFNIYANDPANFDSIRVDVNDVADADLLQIAGDYIKAAGIHN 477
Db 525 AVQAEQLNLHLYIMNIGSILGNDPSANFDGVRIDAVDNVDADLLQIASDYFKEKYRVADN 584
Qy 478 DKAANDHLSILEAWSDDNTPYLDHDDGNMINDNKLRLSLFSLAKPLNQSGMPLTN 537
Db 585 EANAIAHLISILEAWSYNDHQYNKDTKGAQLSIDNPLRETLTTLFLRKNYGRSLSRVITN 644
Qy 538 SLVNRDNDNAETAAPVSPYSPIRAHDSVEQDLIADIIEKAEINPNVVGYSFTMEIEIKAPEI 597
Db 645 SLNRRSEQKHTPRDANIYFVRAHDSVQAVLANIISKQINPKPTDGTFTMDLQKAFEI 704
Qy 598 YNKDLATEKKYTHYNTALSYALLTNKSSVPRVYVYGDFTDDGQYMAHKTINYEAIETL 657
Db 705 YNADIAKADKKTQYNIPAAVATMLTNKDSITRVVYVYGDFTDDGQYMAEKSPYNAIDAL 764
Qy 658 LKARIKYVSGQAMRNQOVGNSEIITSVRYKGAALKATDGTGRTTRTSGVAVIEGNPNSL 717
Db 765 LRARIKYVAGQDMKVTKLNGYEIMSSVRYKGAEEANQLGTAETRNQGMVLTLNRRPDM 824
Qy 718 RLKASDRVVVNMGAHKNQAVRPLLLTTDNGIKAYHSDQEA-AGLVRYTNDRGELIFTA 776
Db 825 KLGANDRLVNMGAHKNQAVRPLLLSKTGLATYTKDSDVPAGLVRYTDQNGNLTFTAD 884
Qy 777 DIKGYANPQVSGYLGVVVPVGAALIKMPALRLARPHQOMASVHQNAALDSRVMPFGFSNF 836
Db 885 DIAHSTVEVSGYLAVVVPVGAASENQDARTKASSTKKEQVFPSSAALDSQVIEGFSNF 944
Qy 837 QAFATKKEEYTNVVIKNDKFAEWGVTDFEMAPQYVSSSTGCSFLDSVIONGYATDRYD 896
Db 945 QDFVKTQSYTNRVIAQNAKLFKEMGITSPFAPQYVSSQDGTFLDSIENGAYAFEDRYD 1004
Qy 897 LGISKPNKYGTADDLVKAIKALHSKGIKVMADWVPDQMYAPPEKEVVTATRVYDKGTPVA 956
Db 1005 IAMSNNKYIGSLKDLMDALRALHAEIGISAIADWVPDQIYNLPKGEVVTASRTNSYGTFRP 1064
Qy 957 GSQIKNTLYVVDGKSSGKQQAKYCGAFLELOAKYPELFAKQISTGVPMDPSPYKIQW 1016
Db 1065 NAEIVNSLYAAKTRTFGNDFOGKYGGAFLELAKYPAIFERVQISNGRKLTTNEKITQW 1124
Qy 1017 SAKYFNGNIIILGRGAYVVLKQOANTYNIISDNKEINFLPKTLNLNQDSQVGSFYDGKGV 1076
Db 1125 SAKYFNGSNIQGTGARVYVLDNATNQYSPSKAGQ--TELPKQ-MTEITGSGFRYDDVQ 1181
Qy 1077 YYSTSGYQAKNTFISEG-DKWWYFDNNGYMTGQSGINGVNYVFLSNGQLRLDALTKNED 1135
Db 1182 YLSYGGYLAKNTFIQVGNQWYTFDKNGNMVTGEQVIDGKYFFLDNGLQLRHVLRQSGD 1241
Qy 1136 GTYAYYNGDGRYENGYYQFMSGVW--RHFN-NGEMSGLTVIDQVQYFDE-MGYQAKG 1191
Db 1242 GHVYVDPKGVQAFNGFYDFAGPRQDVRYFDGNGQMYRGLHDMYGTTFVFDEKTKQAKD 1301
Qy 1192 KFTVTADGKIYFPKQSGMNTNRNPIENEBG-WLYLGEDGAATVTSQTINGCHLYFRAN 1250
Db 1302 KPIRFADGRTYFIPDTGNLAVNRPAQNPENKAWYLDLSNGYAVTGLQTINGKQYFDNE 1361
Qy 1251 GVQVKEFVTDHGRISYSYDNGSGDQIRNRRVNRNAQGWQFYDNNGYAVTGARTINGQLL 1310
```

Db 1362 GROVKGFVTTNNQRY-FLDGDGSEIAPSRFV-TENNKNVYVDNGKLVKGAQVINGNH 1419
Qy 1311 YFRANGVQVKGFEVTDYGRISYDNGSGDQIRNFRVNAQGWFFYFNDNGYAVTGARTI 1370
Db 1420 YFNNDYSQVKGAWNGR-----YDGDGSGQAVSNQFIQIAANQWALNQDGHKVTGLONI 1474
Qy 1371 NGQHLVFRANGVQVKGFEVTDYGRISYDNGSGDQIRNFRVNAQGWFFYFNDNGYAVT 1430
Db 1475 NNKVVYFGSNGAQVKGKLLTVQ-GKCYFDAHTGEQVNRVFEAARGCWYFNSAGAVT 1533
Qy 1431 GARTINGOHLVFRANGVQVKGFEVTDY--GRISYVDANSGE 1470
Db 1534 GOVINGKQLYFDGSGRQVGRIV---YVGKRLFCDAKTGE 1572

RESULT 14
Q8KZL5_9STRE
ID Q8KZL5_9STRE PRELIMINARY; PRT; 1554 AA.
AC Q8KZL5;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Glucosyltransferase.
GN Name-gtfU;
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
RN NCBI_TaxID=1310;
RX [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21958694; PubMed=11960691; DOI=10.1016/S0304-4165(01)00240-9;
RA Hanada N., Fukushima K., Nomura Y., Senpuku H., Hayakawa M.,
RA Mukasa H., Shirota T., Abiko Y.;
RT "Cloning and nucleotide sequence analysis of the Streptococcus
RT sobrinus gtfU gene that produces a highly branched water-soluble
RT glucan.";
RL Biochim. Biophys. Acta 1570:75-79(2002).
DR EMBL; AB089438; BAC07265.1; -; Genomic_DNA.
DR HSSP; P06653; 1H8G.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 5.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1554 AA; 171676 MW; 6981BCC1DAE24A73 CRC64;

Query Match 44.4%; Score 3434.5; DB 2; Length 1554;
Best Local Similarity 46.6%; Pred. No. 1.9e-151;
Matches 731; Conservative 248; Mismatches 458; Indels 131; Gaps 39;

Qy 1 MDKRYKLRKVKRWVTVSVASAVMTLTSLGGLVKAD---SNESKQ----- 46
Db 1 MEKGLHYLKHVKKGHWVTTAVAS-IGLVSLVGAGTVSAEDKRVANDTTAQATVGVDGTGQDQ 59
Qy 47 -ISDNTSVV--TANEESNVITEATSKQEAASQTNHTVTSSSTSVNPKVWNSP 102
Db 60 ATTNDANTTTDTDAQSAN-----TNQDQAGSQDNNQDQAKQDTAN----- 103
Qy 103 YTVGETASNGEKLOQNT-----TVDKTS-----EAAANNISKQTTEADTDVDDNSAAML 153
Db 104 -TDRNQADNSQTDNNQATDQATSPATDGTSVQRDAANVATAADQEGQTAPSEQKSAAL 162
Qy 154 QLEKLPVKEIDGKYIYDNGKVRNTFTLADQKILHFD-ETGAYTDTSDTVNKOIV 212
Db 163 ----SLDNNVKLIDGKYIYVQADGSKYKCNFAITVNGQMLYFSDTSGALSTSYFSQGT 217
Qy 213 TTRSNLYKYNQVDRSAQSFEHVDHLYLTABSWRPKYLKDGKTWTQSTQTEKDFRPLMT 272
Db 218 TNLVDDFSHNKADSTAKSFELNGLTANSYFAGLNRGTQWTEASNENDLRPVLMS 277
Qy 273 WMPDQETQRYVYVNAQLGINKT-YDDTSNQLNLNIAAAATTAQAKIEAKITTLKNTDMLR 331

Db 278 WMPDQETQRYVYVYVNAQVNAQVNAQVNAQVNAQVNAQVNAQVNAQVNAQVNAQVNAQVNA 337
Qy 332 QTIASAVKTKQSAWNSDEKFPD--DHLONGAVLYDNEGKLTTPYANSNYRLARTPTPTNOTG 389
Db 338 TAMEAEVAAQPKNNMSTEN-FNKGDHLOGGALLYTN-SDLTPWANSYRLARTPTPTQDQ 395
Qy 390 KCDPRYADNTIGGVEFLANDVNSNPVVAEQLNHLFNMFGNIYANPDPAFDSIR 449
Db 396 TK--KYFTSGGEGVEFLANDVNSNPVVAEQLNHLFNMFGNIYANPDPAFDSIR 453
Qy 450 VDAVDNVDADLLQIAGDYKAAKGIHKNDKAANDHLSILEAMSDNTPTLHDDGDNNIM 509
Db 454 VDAVDNVDADLLQVSYNFKDNYKVTDSSEANALAHISILEAMSLNDQYNEDTNGTALS 513
Qy 510 DNKRLSLFLSLAKPLNORSQGNPLITNSL-VNRTDDNAETAAPVPSYFIRAHDSVODL 568
Db 514 DNSRLTSLAVITKOPGQIDLSNLISESVNKRANDTAYGDTIPTYSFVRAHDSVQTV 573
Qy 569 IADIIKAEINPNVVGYSFTMEBIKKAFEIYNKDLATEKKYTHYNTALSVALLLTNKSSV 628
Db 574 IAKIVKEIDTNSDGYTFLDQDKAFAKLYNEDMAKVNKYTHYNTIPAAVALLLSNMSV 633
Qy 629 PRVYGDMPDQGYMAHKTINYEAETLLKARIKYVSGQAMRNQOV-GNSBIITSVRY 687
Db 634 PRVYGDLYTDGQYMAKSPYDAITMLQGRYAVVSGQSEVHKVNGNQLSSVRY 693
Qy 688 GKALKATDT-CDRTTRTSGVAVIEGNNPSLRLKASDRVVVNMGAHKKQARPLLLTTD 746
Db 694 GQDLMSADDTQGTDLSTRTSLTVLSNDPNLDEL-GGDSLTVMNGRAHANQARPLILGT 752
Qy 747 NGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKGYANPOVSGVLGVVVPVGAALIKMFAL 806
Db 753 DGVQSYLKDS-D-TNIVKYTDANGNLTFTADDIKGYSTVDMSGYLAVVVPVGA-K--DGQV 809
Qy 807 RLARPHQQMA---SVHQNAALDSRVMEFGFSNFOAPATKKEBYTNVVIARNVDKFAEWG 863
Db 810 RVAADTNQKADGKSLKTSALDSQVIEGFSNFQDFANNADYTNKKAENADFFPKLGI 869
Qy 864 TDFEMAPQVYSTDGSFLDSVLQNGYAFTRDRLGISKPNKYGTADDLVKALKHSGKI 923
Db 870 TSFEMAPQVVSATDGSFLDSIIQNGYAFSDRYDLAMSKNNKYGSKDDLANALKALHANGI 929
Qy 924 KVMADWVPDQMYAFPEKEVVTATRVYDVTGTPVAGSOIKNTLVVDGKSSGKQQAQYGA 983
Db 930 QAIADWVPDQIYQLPGEVVTAKRTNSYGNPTFDAYINNALYATNTKSSGSDYQAQYGA 989
Qy 984 FLUELQAKYFELFARKQISTGVMPDPSVKIKOWSAKYFNGTNILGRGAGYVLKQDQATNTY 1043
Db 990 FLDELKAKYFDMFTVMNISTGKPIDPSTKIKQWEAKYFNGTNVLGKAGYVLSDDATKY 1049
Qy 1044 FNISDNKEINFLPKTLL-NQDSQVGSFVYDGYVYVYSTSYGQAKWTFISEGDKWYFDDN 1102
Db 1050 FTWNEGD--FLPASFTGDQNAKTGYDGTGMAYYSTSGNKAVNSFIYEGGHYFYDKD 1107
Qy 1103 GYMTGTA-QSINGVNYFSLNGLQLRDAILKNEDEGTIAYGNDGRYE-NGYQFM---- 1156
Db 1108 GHMTVGSKAEDGNDGYIFLPNGIQMRDAILYDAQGNSYTYGRTGILYKGNMYPFVDPNN 1167
Qy 1157 --SGYRHFN-NGEMSVGLTVIDGQVQYFDEMGYQAKGKFTVTADGKRYFDFKQSGNMRY 1213
Db 1168 ANKTVRYFDANNVMAIGVRNMYGQTYFYFDEMGYQAKGQLLTDKG-THYFDEDNAMAK 1226
Qy 1214 NRPIENEGKMLYLGEDGAATGSGTINQOHLIFR-ANGVQVKGFEVTDHGRISYDGN 1272
Db 1227 NKVF-NVGDDWYMDGNGNAVKQGPVNNQLIYFNPETGVQVKGQFITDAQGRTSYDAN 1285
Qy 1273 SGQDIRNF-----VRNAQGFYFED-NGYAVTCAITINGOL 1309
Db 1286 SGALKSSGFFTPNGSDWTYAENGYVYKFKQVAENQDQWYIFDQTTGQAKGAARVDGRD 1345
Qy 1310 LYFRA-NGVQVKGFEVTDYGRISYDNGSGDQIRNFRVNAQGWFFYFNDNGYAV---- 1364

THIS PAGE BLANK (USPTO)

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	7143	100.0	1375	5	AAU98028	S. mutans
2	7143	100.0	1375	5	AAU79288	S. streptococcus
3	7143	100.0	1375	7	ADD93555	Streptococcus
4	7143	100.0	1375	9	ADK37278	Streptococcus
5	5392	75.5	1476	5	AAU79284	S. mutans
6	5285	74.0	1475	5	AAU98027	S. mutans
7	5285	74.0	1475	7	ADD93654	Streptococcus
8	5285	74.0	1475	9	ADK37277	Streptococcus
9	5284	74.0	1475	5	AAU98030	S. mutans
10	5281	73.9	1475	5	AAU98040	S. mutans
11	5280	73.9	1475	5	AAU98031	S. mutans
12	5279	73.9	1475	5	AAU98035	S. mutans
13	5279	73.9	1475	5	AAU98033	S. mutans
14	5278	73.9	1475	5	AAU98032	S. mutans
15	5277	73.9	1475	5	AAU98036	S. mutans
16	5273	73.8	1475	5	AAU98034	S. mutans
17	5271	73.8	1475	5	AAU98037	S. mutans
18	5261	73.7	1475	5	AAU98038	S. mutans
19	5258	73.6	1475	5	AAU98039	S. mutans
20	4334.5	60.7	1017	5	AAU79285	Streptococcus
21	4214.5	59.0	1590	7	ADD93657	Streptococcus
22	4214.5	59.0	1590	9	ADK37280	Streptococcus
23	4168.5	58.4	1592	2	AAR32925	Glucosylase
24	3603.5	50.4	1430	5	AAU98044	S. mutans

CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
CC GTF mutant, an isolated polynucleotide which encodes F1 or F2, or its
CC complementary polynucleotide, a ribonucleic acid sequence encoding the
CC GTF mutant, an expression cassette comprising the polynucleotide operably
CC linked to a promoter, a vector comprising the expression cassette, host
CC cell introduced with the vector, a transgenic plant comprising the
CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
CC coating composition comprising a glucan produced in a plant transformed
CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
CC thermoplastic molecule or their combinations or glucan and starch where
CC the glucan is produced in the amyloplast and/or vacuole or a maize line
CC deficient in starch biosynthesis, transformed with a gene encoding a
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
CC comprising the glucan (paper sizing/coating agent). The vector is useful
CC for producing a glucan in a plant. The method comprises transforming a
CC plant cell with the vector, growing the plant cell under plant growing
CC conditions to produce a regenerated plant and inducing expression of the
CC polynucleotide for a time sufficient to produce the glucan in the
CC regenerated plant, where the vector contains a transit sequence from
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
CC is produced in the amyloplast of potato or the vacuole of sugar beet.
CC Glucans are useful as substitutes for and additions to modified starch
CC and latexes in paper manufacture. Unlike prior art techniques, which
CC require input materials that produce chemical effluents, paper
CC manufacture utilising the glucan produced by GTF, which utilises
CC biologically produced input materials, is more cost-effective and
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
CC properties and impart gloss to the paper during coating step. The present
CC sequence represents GTFC
XX
SQ Sequence 1375 AA;

Query Match 100.0%; Score 7143; DB 5; Length 1375;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MEKVRFLKRVKRWVTSIASAVVTLTSLGSLVKADSTDDRQQAATESQASLVTTSE 60
DB 1 MEKVRFLKRVKRWVTSIASAVVTLTSLGSLVKADSTDDRQQAATESQASLVTTSE 60
QY 61 AAKETLTATDTSTATSAISQPTATVDNVSTTNQSTNTTANTANFVVKPTTSEAKTDN 120
DB 61 AAKETLTATDTSTATSAISQPTATVDNVSTTNQSTNTTANTANFVVKPTTSEAKTDN 120
QY 121 SDKIITTSKAVNRLTATGKFPANNNTAHPKTVTDK1VIPKIGKLGKQPSLSQDDIAA 180
DB 121 SDKIITTSKAVNRLTATGKFPANNNTAHPKTVTDK1VIPKIGKLGKQPSLSQDDIAA 180
QY 181 LGNVQNRKNGKYYKEDGTLQKNYALNINGKTFPFDETGALSNNVTLPSKKGNITND 240
DB 181 LGNVQNRKNGKYYKEDGTLQKNYALNINGKTFPFDETGALSNNVTLPSKKGNITND 240
QY 241 NTNSPAQNOVYSTDVANFEHVDHYLTAESWTRPKYILKDGKTWTQSTEKDFRPLMTTW 300
DB 241 NTNSPAQNOVYSTDVANFEHVDHYLTAESWTRPKYILKDGKTWTQSTEKDFRPLMTTW 300
QY 301 PDQETQROYVNMNAQLGHOTYNTATSPQLNLAAQTIQTKIEKITAENKTNWLRQTI 360
DB 301 PDQETQROYVNMNAQLGHOTYNTATSPQLNLAAQTIQTKIEKITAENKTNWLRQTI 360
QY 361 SAFVKTQSAWNSDSEKPDHDLQKALLYSNNSKULTSOANSNYRILNRTPTNQTGKQDPR 420
DB 361 SAFVKTQSAWNSDSEKPDHDLQKALLYSNNSKULTSOANSNYRILNRTPTNQTGKQDPR 420
QY 421 YTADRTIGYEFLANDVNSNPVQAOLNWLHFLMNFNIYANDPDANFDSIRVDVAD 480
DB 421 YTADRTIGYEFLANDVNSNPVQAOLNWLHFLMNFNIYANDPDANFDSIRVDVAD 480
QY 481 NVADLLQIAGDYLKAAKGIHKNDHLSILEAWSYNPTPYLHDDGDNMINDNRLR 540
DB 481 NVADLLQIAGDYLKAAKGIHKNDHLSILEAWSYNPTPYLHDDGDNMINDNRLR 540

QY 541 LSLYSLAKPLNQRSGMNPILTNSLVNRTDDNAETAAVPSYFTRAHDSEVQDLIRNIIR 600
DB 541 LSLYSLAKPLNQRSGMNPILTNSLVNRTDDNAETAAVPSYFTRAHDSEVQDLIRNIIR 600
QY 601 TEINPNVGVSYFTTEEEKAFEIYNKOLLATEKKYTHYNTALSVALLLTNKSSVPRVYVG 660
DB 601 TEINPNVGVSYFTTEEEKAFEIYNKOLLATEKKYTHYNTALSVALLLTNKSSVPRVYVG 660
QY 661 DMFTDDGQYMAHKTINYEAIETLLKARIKYVSGGQAMENQQVGNSEIITTSVRYGKALKA 720
DB 661 DMFTDDGQYMAHKTINYEAIETLLKARIKYVSGGQAMENQQVGNSEIITTSVRYGKALKA 720
QY 721 TDTGDRTRTSGVAVIEGNNPSLRLKASDRVVVNGAAHKNQAVRPLLLTTDNGIKAYHS 780
DB 721 TDTGDRTRTSGVAVIEGNNPSLRLKASDRVVVNGAAHKNQAVRPLLLTTDNGIKAYHS 780
QY 781 DQEAAGLVRYTNDRGELIFTAADIKGYANPOVSGYLVGVVPGAAADQDVRVAASTAPST 840
DB 781 DQEAAGLVRYTNDRGELIFTAADIKGYANPOVSGYLVGVVPGAAADQDVRVAASTAPST 840
QY 841 DGKSVHQNAALDSRVMEFGFSNFQAFATKBEYTNVVIKXNDKPAEAGVTDFFEMAPQVY 900
DB 841 DGKSVHQNAALDSRVMEFGFSNFQAFATKBEYTNVVIKXNDKPAEAGVTDFFEMAPQVY 900
QY 901 SSTGGSFLDSVIQNGYAFTRDYDLGISKPNKYGTADDLVKAIKALHSGIKVMADWVPDQ 960
DB 901 SSTGGSFLDSVIQNGYAFTRDYDLGISKPNKYGTADDLVKAIKALHSGIKVMADWVPDQ 960
QY 961 MYALPEKEVTVATRDVKYGTVPAGSQIKNTLYVVDGSSGKQQAQKYGAFLEELQAKYP 1020
DB 961 MYALPEKEVTVATRDVKYGTVPAGSQIKNTLYVVDGSSGKQQAQKYGAFLEELQAKYP 1020
QY 1021 ELFPARKQISTGVPMDPSVKIKQWSAKYPNGTNILGRGAGYVLKQQAATNTYFSLVSDNTFL 1080
DB 1021 ELFPARKQISTGVPMDPSVKIKQWSAKYPNGTNILGRGAGYVLKQQAATNTYFSLVSDNTFL 1080
QY 1081 PKSLVNPNGHTSSSVTGLVFDGKGYVYSTSGNOAKNAFISLGNWYVFDNNGYVMTGAQ 1140
DB 1081 PKSLVNPNGHTSSSVTGLVFDGKGYVYSTSGNOAKNAFISLGNWYVFDNNGYVMTGAQ 1140
QY 1141 SINGANYFYLSNGIQLRNAIYDNGNKVLSYDNGDRRYENGYYLFGQOWRYFQNGIMAVG 1200
DB 1141 SINGANYFYLSNGIQLRNAIYDNGNKVLSYDNGDRRYENGYYLFGQOWRYFQNGIMAVG 1200
QY 1201 LTRVHGAVQYVDSGFOAKGQFIITADGKLYRFDKSGNQISNRPVRNSKGEWFLFDHNG 1260
DB 1201 LTRVHGAVQYVDSGFOAKGQFIITADGKLYRFDKSGNQISNRPVRNSKGEWFLFDHNG 1260
QY 1261 VAVTGTVTENGORLYFKPENGVOAKGEFIRDANGYLRYVDNPSGNEVRNRFVRNSKGEWFL 1320
DB 1261 VAVTGTVTENGORLYFKPENGVOAKGEFIRDANGYLRYVDNPSGNEVRNRFVRNSKGEWFL 1320
QY 1321 FDHNGIATGARVNVNGHASILSLMVFRLRESSLSQSVKVSNTMILIPMKFVIVM 1375
DB 1321 FDHNGIATGARVNVNGHASILSLMVFRLRESSLSQSVKVSNTMILIPMKFVIVM 1375
RESULT 2
AAU79288
ID AAU79288 standard; protein; 1375 AA.
XX AAU79288;
AC AAU79288;
XX
DT 13-AUG-2002 (first entry)
XX Streptococcus mutans monoclonal antibody-related protein #5.
XX Antibody; dental caries; water insoluble glucan synthetase; anti-caries;
KW glucosyl transferase-B; immunotherapy.
XX
OS Streptococcus mutans.
XX

PN JP2002114709-A.
 XX
 PD 16-APR-2002.
 XX
 PF 04-OCT-2000; 2000JP-00304889.
 XX
 PR 04-OCT-2000; 2000JP-00304889.
 XX
 PA (UYNI-) UNIV NIPPON.
 XX
 XX WPI; 2002-448885/48.
 DR
 XX
 XX Anti-carries agent composed of a monoclonal antibody against an inhibitory
 PT enzyme against water insoluble glucan synthetase of glucosyl transferase-
 PT B (GTF-B) of Streptococcus mutans.
 XX
 XX Disclosure; Page 22-25; 28pp; Japanese.
 PS
 XX The invention relates to a monoclonal antibody against dental caries and
 CC an anti-carries agent composed of a monoclonal antibody produced by
 CC Streptococcus mutans, particularly mouse-hybridoma MHP126 (FERM P-17566)
 CC or mouse-hybridoma MHP136 (FERM P-17567), against an enzyme having
 CC inhibitive activity against water insoluble glucan synthetase of glucosyl
 CC transferase-B. The monoclonal antibody specifically inhibits water
 CC insoluble glucan synthetase of Streptococcus mutans produced glucosyl
 CC transferase-B and is used in the immunotherapy of dental caries. This
 CC sequence represents a Streptococcus mutans monoclonal antibody-related
 CC protein
 XX
 XX Sequence 1375 AA;
 SQ

Query Match 100.0%; Score 7143; DB 5; Length 1375;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEKVRFLKRVKRWVTSIASAVVTLTSLGSLVKADSTDDRQAVTESQASLVTTSE 60
 DB 1 MEKVRFLKRVKRWVTSIASAVVTLTSLGSLVKADSTDDRQAVTESQASLVTTSE 60
 QY 61 AAKETLTATDTSTATSQPTATVTDNVSTTNQSTNTTANTANFVVKPTTTSQAKTDN 120
 DB 61 AAKETLTATDTSTATSQPTATVTDNVSTTNQSTNTTANTANFVVKPTTTSQAKTDN 120
 QY 121 SDKIITTSKAVNRLTATGKVFVANNNTAHPKVTVDKIVPKIKGLKQPSLSQDDIAA 180
 DB 121 SDKIITTSKAVNRLTATGKVFVANNNTAHPKVTVDKIVPKIKGLKQPSLSQDDIAA 180
 QY 181 LGNVKNIKRVNGKYTYKEDGTLOKNYALNINNGKTPFPDETGALSNNLPSKKGNIITND 240
 DB 181 LGNVKNIKRVNGKYTYKEDGTLOKNYALNINNGKTPFPDETGALSNNLPSKKGNIITND 240
 QY 241 NTNSFAQNOVYSTDVANFEVDHYLTAEWSYRPKYILKDGKTWTQSTEKDFRPLMTTW 300
 DB 241 NTNSFAQNOVYSTDVANFEVDHYLTAEWSYRPKYILKDGKTWTQSTEKDFRPLMTTW 300
 QY 301 PQDETQROVYVYVNAQLGTHQYNTATSPQLNLAAQITQTKIEBKITAENKTNWLRQTI 360
 DB 301 PQDETQROVYVYVNAQLGTHQYNTATSPQLNLAAQITQTKIEBKITAENKTNWLRQTI 360
 QY 361 SAFVKTQSAWNSDSKPFDDHLQKALLYSNNSKLTSQANSNYRILNRTPTNQTCKDPR 420
 DB 361 SAFVKTQSAWNSDSKPFDDHLQKALLYSNNSKLTSQANSNYRILNRTPTNQTCKDPR 420
 QY 421 YTADRTIGGYEFLANDVNSPVVQAEQLNHLFNMFGNIYANDPDANFDSIRDAVD 480
 DB 421 YTADRTIGGYEFLANDVNSPVVQAEQLNHLFNMFGNIYANDPDANFDSIRDAVD 480
 QY 481 NVADLLQIAGDYKAAAGIHKNDKAAANDHLSLEAWSYNDTPYLHDDGDMNMDNRLR 540
 DB 481 NVADLLQIAGDYKAAAGIHKNDKAAANDHLSLEAWSYNDTPYLHDDGDMNMDNRLR 540
 QY 541 LSLYSLAKPLNQRSGMPLITNSLVNRTDDNAETAAPVPSYFIRAHDSEVQDLIRNIIR 600
 DB 541 LSLYSLAKPLNQRSGMPLITNSLVNRTDDNAETAAPVPSYFIRAHDSEVQDLIRNIIR 600

DB 541 LSLYSLAKPLNQRSGMPLITNSLVNRTDDNAETAAPVPSYFIRAHDSEVQDLIRNIIR 600
 QY 601 TEINENVVGYSTTBEIKKAFIYNKDLATEKKYTHYNTALSYALLTNKSKSVPRVYG 660
 DB 601 TEINENVVGYSTTBEIKKAFIYNKDLATEKKYTHYNTALSYALLTNKSKSVPRVYG 660
 QY 661 DMFTDDGQYMAHKTINYEAEITLLKARIKYVSGGQAMRNQOVGNSIITSVRYGKALK 720
 DB 661 DMFTDDGQYMAHKTINYEAEITLLKARIKYVSGGQAMRNQOVGNSIITSVRYGKALK 720
 QY 721 TDTGRTTTSVAVIEGNNPSLRKASDRVVVNNGAHKNQAYRPLLLITTDNGIKAYHS 780
 DB 721 TDTGRTTTSVAVIEGNNPSLRKASDRVVVNNGAHKNQAYRPLLLITTDNGIKAYHS 780
 QY 781 DQEAAGLVRYTNDRGELIFTAADIKGYANPQVSGVLGVWVPVGAADQDVRVAASAPST 840
 DB 781 DQEAAGLVRYTNDRGELIFTAADIKGYANPQVSGVLGVWVPVGAADQDVRVAASAPST 840
 QY 841 DGKSVHQAALDSRVVMFSGFSNFQAFATKKEBYTNVVIKQVDFKFAEWGVTDFEMAPQV 900
 DB 841 DGKSVHQAALDSRVVMFSGFSNFQAFATKKEBYTNVVIKQVDFKFAEWGVTDFEMAPQV 900
 QY 901 SSTDSFLDSVIQNGYAFTRDYLIGISKPNKYGTADDLVKAIKALHSKGIKVMADWVPDQ 960
 DB 901 SSTDSFLDSVIQNGYAFTRDYLIGISKPNKYGTADDLVKAIKALHSKGIKVMADWVPDQ 960
 QY 961 MYALPEKEVWVTRVDKYGTPVAGSQIKNTLVVDGKSGKQDQAKYGGAFLEELQAKYP 1020
 DB 961 MYALPEKEVWVTRVDKYGTPVAGSQIKNTLVVDGKSGKQDQAKYGGAFLEELQAKYP 1020
 QY 1021 ELFAKQISTGVPMDPSVKIKOWSAKYFNGTNILGRGAGYVLKDQATNTYFSLVSDNTFL 1080
 DB 1021 ELFAKQISTGVPMDPSVKIKOWSAKYFNGTNILGRGAGYVLKDQATNTYFSLVSDNTFL 1080
 QY 1081 PKSLVNPNGHTSSSVTGLVFDGKGYVYVYSTSGNQAKNAFISLGNWYFPDNNGYMVTGAQ 1140
 DB 1081 PKSLVNPNGHTSSSVTGLVFDGKGYVYVYSTSGNQAKNAFISLGNWYFPDNNGYMVTGAQ 1140
 QY 1141 SINGANYYPLSNGIQLRNAIYDNGNKVLSYNGDGRYENGYYLFGQWRYPQNGIMAVG 1200
 DB 1141 SINGANYYPLSNGIQLRNAIYDNGNKVLSYNGDGRYENGYYLFGQWRYPQNGIMAVG 1200
 QY 1201 LTRVHGAQVYFDASGFOAKGQFITTADGKLRYPDRDSGNQISNRRFVRNSKGEWFLPDHNG 1260
 DB 1201 LTRVHGAQVYFDASGFOAKGQFITTADGKLRYPDRDSGNQISNRRFVRNSKGEWFLPDHNG 1260
 QY 1261 VAVTGTVTTFNGORLYFKPENGVOAKGEPTRDANGYLRYVDPNSGNEVRNRFVRNSKGEWFL 1320
 DB 1261 VAVTGTVTTFNGORLYFKPENGVOAKGEPTRDANGYLRYVDPNSGNEVRNRFVRNSKGEWFL 1320
 QY 1321 FDHNGIATVGARVNVGHASIIISLMVFRLESLSQSVKVVSNMTILIPEMKFVIM 1375
 DB 1321 FDHNGIATVGARVNVGHASIIISLMVFRLESLSQSVKVVSNMTILIPEMKFVIM 1375

RESULT 3
 ADD93655
 ID ADD93655 standard; protein; 1375 AA.
 XX
 AC ADD93655;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Streptococcus mutans glucosyltransferase-C.
 XX
 KW Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.
 XX
 OS Streptococcus mutans.
 XX
 FN WO2003075845-A2.
 XX
 PD 18-SEP-2003.
 XX

PD 10-FEB-2005.
XX
PF 09-MAR-2004; 2004US-00797821.
XX
PR 13-APR-1998; 98US-0081550P.
PR 08-JAN-1999; 99US-0115142P.
PR 12-APR-1999; 99US-00290049.
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
PR 07-MAR-2003; 2003US-00383930.
XX
PA (SMIT/) SMITH D J.
PA (TAUB/) TAUBMAN M A.
XX
XX Smith DJ, Taubman MA;
XX WPI; 2005-151644/16.
XX
XX New composition comprising a fragment of a glucan binding protein-B
PT (GbpB) that binds to MHC class II protein, and a biocompatible
PT microparticle, useful for producing an antibody (claimed) for immunizing
PT mammals against dental caries.
XX
XX Claim 7; SEQ ID NO 35; 73pp; English.
XX
XX The invention relates to a composition comprising a fragment of a glucan
CC binding protein-B (GbpB) and a biocompatible microparticle, where the
CC fragment binds to a major histocompatibility complex (MHC) class II
CC protein. The composition is useful for producing an antibody for
CC immunizing mammals against dental caries. This sequence corresponds to a
CC Streptococcus mutans GbpB protein of the invention.
XX
XX Sequence 1375 AA;
Query Match 100.0%; Score 7143; DB 9; Length 1375;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 MEKKVRFKLKVKRWVTVSIAVVTLSLGSGLVKADSTDDRQAVTESQASLVTTS 60
1 MEKKVRFKLKVKRWVTVSIAVVTLSLGSGLVKADSTDDRQAVTESQASLVTTS 60
61 AAKETLTATDTSTATSQPTATVDNVSTTNQSTNTTANTANFVVKPTTSEQAKTN 120
61 AAKETLTATDTSTATSQPTATVDNVSTTNQSTNTTANTANFVVKPTTSEQAKTN 120
121 SDKIITTSKAVNRLTATGKRVFVANNNTAHPKVTVDKIPIKIGKLKQPSLSQDDIAA 180
121 SDKIITTSKAVNRLTATGKRVFVANNNTAHPKVTVDKIPIKIGKLKQPSLSQDDIAA 180
181 LGNVNIRKVNKGYYYKEDGTLOKNYALNINNGKTFPFDETGALSNTLPSKKGNIITND 240
181 LGNVNIRKVNKGYYYKEDGTLOKNYALNINNGKTFPFDETGALSNTLPSKKGNIITND 240
241 NTNSFAQVNVYSTVDVAFVHDVHLYTAESWYRPKYILKDGKTTQSTEKDFRPLMTWM 300
241 NTNSFAQVNVYSTVDVAFVHDVHLYTAESWYRPKYILKDGKTTQSTEKDFRPLMTWM 300
301 PQDETQROVYVYVNNALQGLHQTNTATSPLOLNLAQAQTIQTKEIKITAEKNTNMLRQTI 360
301 PQDETQROVYVYVNNALQGLHQTNTATSPLOLNLAQAQTIQTKEIKITAEKNTNMLRQTI 360
361 SAFVKTQSAWNSDSKPPDDHLQKGLLYSNNKSLTQSAWNSYRILNRTPTNQTCKKOPR 420
361 SAFVKTQSAWNSDSKPPDDHLQKGLLYSNNKSLTQSAWNSYRILNRTPTNQTCKKOPR 420
421 YTADRTIGGYEFLANDVNSNPVQAEQLNWLHFLMFGNITYANDPDANFDSIRVDAVD 480
421 YTADRTIGGYEFLANDVNSNPVQAEQLNWLHFLMFGNITYANDPDANFDSIRVDAVD 480
481 NYDADLLOIAGYLYLKAAGIHKNDKKAANDHLSILEAWSYNDTPPYLHDDGDNMNDNRLR 540
481 NYDADLLOIAGYLYLKAAGIHKNDKKAANDHLSILEAWSYNDTPPYLHDDGDNMNDNRLR 540

QY 541 LSLLYSLAKPLNORSQMNPLITNSLVNRTDDNAETAAPSYSFIRAHDSVQDLIRNIIR 600
DB 541 LSLLYSLAKPLNORSQMNPLITNSLVNRTDDNAETAAPSYSFIRAHDSVQDLIRNIIR 600
QY 601 TEINPNVGYSTFTEBIKKAFIYNKDLATEKKYTHYNTALSYALLTNKSSVPRVYV 660
DB 601 TEINPNVGYSTFTEBIKKAFIYNKDLATEKKYTHYNTALSYALLTNKSSVPRVYV 660
QY 661 DMFTDDGOVMAKNTINYEAETLLKARIKYVSGQAMRNQOVNSIIITSVRVKGKALKA 720
DB 661 DMFTDDGOVMAKNTINYEAETLLKARIKYVSGQAMRNQOVNSIIITSVRVKGKALKA 720
QY 721 TDTGRTTSTGVAVIEGNNPSLRKASDRVVVNMGAHKNQAYRPLLLTTDNGIKAYHS 780
DB 721 TDTGRTTSTGVAVIEGNNPSLRKASDRVVVNMGAHKNQAYRPLLLTTDNGIKAYHS 780
QY 781 DQEAAGLVRYTNDRGELIFTAADIKGYANPQVSGYLGVMVPGVGAADQDVRVAASTAPST 840
DB 781 DQEAAGLVRYTNDRGELIFTAADIKGYANPQVSGYLGVMVPGVGAADQDVRVAASTAPST 840
QY 841 DGKSVHQNAAALDSRVMEFGFSNFQAFATKKEBYTNVVIKQVNDKFAEWGVTDFEMAPOVY 900
DB 841 DGKSVHQNAAALDSRVMEFGFSNFQAFATKKEBYTNVVIKQVNDKFAEWGVTDFEMAPOVY 900
QY 901 SSTDGSFLDSVTONGYAPFTDRYDLGIGSKPKYGTADDLVKAIKALHSKGIKVMADWVPDQ 960
DB 901 SSTDGSFLDSVTONGYAPFTDRYDLGIGSKPKYGTADDLVKAIKALHSKGIKVMADWVPDQ 960
QY 961 MYALPEKEVVTATRVVDKYGTPVAGSQIKNTLVVVDGKSSGKQDQAKYCGAFLEELQAKYP 1020
DB 961 MYALPEKEVVTATRVVDKYGTPVAGSQIKNTLVVVDGKSSGKQDQAKYCGAFLEELQAKYP 1020
QY 1021 ELFARQKQISTGVPMDSVKIKQMSAKYFNGTNILRGAGYVLKQDQATNTYFSLVSDNTFL 1080
DB 1021 ELFARQKQISTGVPMDSVKIKQMSAKYFNGTNILRGAGYVLKQDQATNTYFSLVSDNTFL 1080
QY 1081 PKSLVNPNGHSTSSVTGLVDFDGKGVYVYSTSGNQAKNAFISLGNWYFDNNGYMVTGAQ 1140
DB 1081 PKSLVNPNGHSTSSVTGLVDFDGKGVYVYSTSGNQAKNAFISLGNWYFDNNGYMVTGAQ 1140
QY 1141 SINGANYFVLSNGIQLRNAIYDNGNKVLSYYGNDGRRYENGYYLFGQWRIFYQNGIMAVG 1200
DB 1141 SINGANYFVLSNGIQLRNAIYDNGNKVLSYYGNDGRRYENGYYLFGQWRIFYQNGIMAVG 1200
QY 1201 LTRVHGAVQYFDASGFOAKGQFIITADGKLYFDRDSGNQISNRFVNRNSKGWFLPDHNG 1260
DB 1201 LTRVHGAVQYFDASGFOAKGQFIITADGKLYFDRDSGNQISNRFVNRNSKGWFLPDHNG 1260
QY 1261 VAVTGTVTFTNGORLYFKPENGVOAKGEFIRDANGYLRYPDPSGNEVRNRFVNRNSKGWFL 1320
DB 1261 VAVTGTVTFTNGORLYFKPENGVOAKGEFIRDANGYLRYPDPSGNEVRNRFVNRNSKGWFL 1320
QY 1321 FDHNGIAGTVGARVNVNGHASIISLWVFLRESSLSQSVKVSNTMILIPKFFVIM 1375
DB 1321 FDHNGIAGTVGARVNVNGHASIISLWVFLRESSLSQSVKVSNTMILIPKFFVIM 1375

RESULT 5
AAU79284
ID AAU79284 standard; protein; 1476 AA.
XX
AC AAU79284;
XX
DT 13-AUG-2002 (first entry)
XX
DE Streptococcus mutans monoclonal antibody-related protein #1.
XX
KW Antibody; dental caries; water insoluble glucan synthetase; anti-carries;
KW glucosyl transferase-B; immunotherapy.
XX
OS Streptococcus mutans.
XX

Db 509 MDNKLRLSLFLSLAKPLNQRSGMNPILITNSLVNRTDDNAETAAPSPYSPIRADHSEVQDL 568
QY 595 IRIINIRTEINPNVVGYSFTTEIKKAFEIYNKDLATEKKYTHYNTALSVALLLTNKSSV 654
Db 569 IADIIKAEINPNVVGYSFTTEIKKAFEIYNKDLATEKKYTHYNTALSVALLLTNKSSV 628
QY 655 PRVYGDMDTDDGQYMAHKTINYEAIETLLKARIKYVSGQAMRNQOVGNSEIITSVRYG 714
Db 629 PRVYGDMDTDDGQYMAHKTINYEAIETLLKARIKYVSGQAMRNQOVGNSEIITSVRYG 688
QY 715 KGALKATDGTGTRTSGVAVTEGNPNLSRLKASDRVVVNMGAHKNAQYRPLLLITDNG 774
Db 689 KGALKATDGTGTRTSGVAVTEGNPNLSRLKASDRVVVNMGAHKNAQYRPLLLITDNG 748
QY 775 IKAYHSDQEAAGLVRYTNDRGELIPTAADIKGYANPQVSGYLGWVPVCGAAADQ--DVRV 832
Db 749 IKAYHSDQEAAGLVRYTNDRGELIPTAADIKGYANPQVSGYLGWVPVCGAALIKFALKL 808
QY 833 AASTAPSTDGKSHVQNAALDSRVPEGFNSFOAFATKKEEYTNVVIKNDVKFAEWGVTD 892
Db 809 A--RPHQQWASVHQNAALDSRVPEGFNSFOAFATKKEEYTNVVIKNDVKFAEWGVTD 865
QY 893 FEMAPQYVSSDGSFLDSVIONGYAFTDRYDLGISKPNKYGTADDLVKAIKALHSGKIKV 952
Db 866 FEMAPQYVSSDGSFLDSVIONGYAFTDRYDLGISKPNKYGTADDLVKAIKALHSGKIKV 925
QY 953 MADWVPDQMYALPEKEVVTATRVKYGTPVAGSQIKNTLYVVDGSSGKQOAKYGGAPL 1012
Db 926 MADWVPDQMYALPEKEVVTATRVKYGTPVAGSQIKNTLYVVDGSSGKQOAKYGGAPL 985
QY 1013 BELQAKYPELFARKQISTGVPMDSVVKIKQWSAKYFNGTNIILGRGAGYVLKQDQATNTVFS 1072
Db 986 BELQAKYPELFARKQISTGVPMDSVVKIKQWSAKYFNGTNIILGRGAGYVLKQDQATNTVFS 1045
QY 1073 LVSDN---TFLPKSLWPNHGTSSVTGLVFDGKGVYVYSTSGNQAQNAFISLGNWYVF 1129
Db 1046 -ISDNKEINFLPKTLN-----QDSQVGFSDYDGKGVYVYSTSGYQAKNTFFISEGDKWYVF 1099
QY 1130 DNNGYVMTGAQSGINGANYFLSNGQLRINAIYDNGKNVLSYYGNDGRYENGYYLFP-CQQ 1188
Db 1100 DNNGYVMTGAQSGINGANYFLSNGQLRINAIYDNGKNVLSYYGNDGRYENGYYLFP-CQQ 1159
QY 1189 WRYFQNGIMAVGLTRVHGAVQYFDASGFOAKGQFITTADGKRLYFDRDSGNQISNRVFN 1248
Db 1160 WEHFNNGEMSVGLTVIDGQVQYFDEMGYQAKGFVTTADGKRYFDKQSGNMYRNFEN 1219
QY 1249 SKGEWFLFDHNGVAVTGVTFNGQRLYFKPVGQVQAKGFEIRDANGYLRYDPNSGNEVRN 1308
Db 1220 BEGKWLXLGDEGAAVTSQTNGQHLVFRANGVQVKGFEVTDHGRISYDNGSGDQIRN 1279
QY 1309 RFRNRSKGEWFLFDHNGVAVTGVTFNGQRLYFKPVGQVQAKGFEIRDANGYLRYDPNSGNEVRN 1347
Db 1280 RFRNRSKGEWFLFDHNGVAVTGVTFNGQRLYFKPVGQVQAKGFEIRDANGYLRYDPNSGNEVRN 1313

RESULT 10

ID AAU98040
XX AAU98040 standard; protein; 1475 AA.
AC AAU98040;
XX
DT 27-AUG-2002 (first entry)
XX
DE S. mutans glucosyltransferase GTFB mutant K779Q.
KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture; mutant; mutin.
XX
OS Streptococcus mutans.
XX Synthetic.

Key Location/Qualifiers
Misc-difference 779
/note= "Wild-type Lys substituted by Gln"

US2002031826-A1.

14-MAR-2002.

19-DEC-2000; 2000US-00740274.

07-JUN-1995; 95US-00478704.

07-JUN-1995; 95US-00482711.

16-JAN-1998; 98US-00007999.

16-JAN-1998; 98US-00008172.

20-JAN-1998; 98US-00009620.

11-DEC-1998; 98US-00210361.

(NICH/) NICHOLS S E.

Nichols SE;

WPI; 2002-414332/44.

Glucosyltransferase B or D protein useful for producing a glucan useful as substitutes for and additions to modified starch and latexes in paper manufacture, comprises mutations in specific positions.

Claim 36; Page; 44pp; English.

The invention an isolated protein comprising a glucosyltransferase (GTF) B polypeptide having changes at position from 1448V, D457N, D567T, K1014T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T, I448V/D457N/D567T/D571K/K1014T, Y169A/Y170A/Y171A, and K779Q or a GTF D polypeptide having changes at positions from T598D, T598E, N471D, N471D/T598D, and N471D/T598E. Also included are a glucan produced by the GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its complementary polynucleotide, a ribonucleic acid sequence encoding the GTF mutant, an expression cassette comprising the polynucleotide operably linked to a promoter, a vector comprising the expression cassette, host cell introduced with the vector, a transgenic plant comprising the vector, a seed or tuber from the transgenic plant, a paper sizing and/or coating composition comprising a glucan produced in a plant transformed with a gene encoding the mutant GTF, wild type or, starch, a latex, thermoplastic molecule or their combinations or glucan and starch where the glucan is produced in the amyloplast and/or vacuole or a maize line deficient in starch biosynthesis, transformed with a gene encoding a glucosyltransferase B or D enzyme, wild-type or mutant and a paper comprising the glucan (paper sizing/coating agent). The vector is useful for producing a glucan in a plant. The method comprises transforming a plant cell with the vector, growing the plant cell under plant growing conditions to produce a regenerated plant and inducing expression of the polynucleotide for a time sufficient to produce the glucan in the regenerated plant, where the vector contains a transit sequence from ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and chlorophyll AB binding protein to produce a transgenic plant, and glucan is produced in the amyloplast of potato or the vacuole of sugar beet. Glucans are useful as substitutes for and additions to modified starch and latexes in paper manufacture. Unlike prior art techniques, which require input materials that produce chemical effluents, paper manufacture utilizing the glucan produced by GTF, which utilizes biologically produced input materials, is more cost-effective and environmentally friendly. Moreover, glucans also exhibit thermoplastic properties and impart gloss to the paper during coating step. The present sequence represents a GTFB mutant of the invention. Note: The present sequence is not shown in the specification but was created by the indexer using the GTFB sequence appearing as AAU98027 and the information in claim 36

Sequence 1475 AA;

Query Match 73.9%; Score 5281; DB 5; Length 1475;
Best Local Similarity 75.8%; Pred. No. 0;

Db 329 WLRQTISAFVKTSQAWNSDEKPFDDHLQNGAVLYDNEGKLTTPYANSNYRIILNRPPTNQ 388
Qy 415 GKDPRTYADRTIGGYEFLANDVNSVPVQAEQLNLWLFMLNFGNIYANDPDANFOSI 474
Db 389 GKDPRTYADNTIGGYEFLANDVNSVPVQAEQLNLWLFMLNFGNIYANDPDANFOSI 448
Qy 475 RYDAVNDVADLLQTAGDYLYKAAKGIHKNDKKAANDHLSILEAWSYNDTPYLLHDDGDNMIN 534
Db 449 RYDAVNDVADLLQTAGDYLYKAAKGIHKNDKKAANDHLSILEAWSYNDTPYLLHDDGDNMIN 508
Qy 535 MDNRLSLLYSLAKPLNORSQMNPLITNSLVNRTDDNAETAAPVPSYFIRAHDSEVDL 594
Db 509 MDNKLRLSLFLSLAKPLNORSQMNPLITNSLVNRTDDNAETAAPVPSYFIRAHDSEVDL 568
Qy 595 IRNIIRTEINPNVVGYSFTTEBEIKKAFIYKNDLLATEKKYTHYNTALSYALLTNKSSV 654
Db 569 IADIIFKABEINPNVVGYSFTTEBEIKKAFIYKNDLLATEKKYTHYNTALSYALLTNKSSV 628
Qy 655 PRVYVGMFTDDGQYMAHKTINYEAIETLLKARIKYVSGGQAMRNQOVGNSEIITSVRYG 714
Db 629 PRVYVGMFTDDGQYMAHKTINYEAIETLLKARIKYVSGGQAMRNQOVGNSEIITSVRYG 688
Qy 715 KGALKATDTGRTTTSVAVIEGNNPSRLKASDRVVVNNGAAHKNQAYRPLLLTTDNG 774
Db 689 KGALKATDTGRTTTSVAVIEGNNPSRLKASDRVVVNNGAAHKNQAYRPLLLTTDNG 748
Qy 775 IKAYHSDQEAAGLVRYTDRGELIFTAADIKGYANPQVSGYLVGVVVPVGAADQ--DVRV 832
Db 749 IKAYHSDQEAAGLVRYTDRGELIFTAADIKGYANPQVSGYLVGVVVPVGAALIKMFALRL 808
Qy 833 AASTARSTDGKSVHQAALDSRWMEFGSNFOAFATKKEEYTNVVIKQVDFKFAEWGTYD 892
Db 809 A---RPHQOMASVHQAALDSRWMEFGSNFOAFATKKEEYTNVVIKQVDFKFAEWGTYD 865
Qy 893 FEMAPQYVSSDGSFLDSVIONGYAFTDRYDLGISKPNKYGTADDLVKAIKALHSGKIKV 952
Db 866 FEMAPQYVSSDGSFLDSVIONGYAFTDRYDLGISKPNKYGTADDLVKAIKALHSGKIKV 925
Qy 953 MADWVPDQMYALPEKEVVTATRVDKYTPVAGSQIKNTLYYVDGKSGKQOQAKYGGAPL 1012
Db 926 MADWVPDQMYAPPEKEVVTATRVDKYTPVAGSQIKNTLYYVDGKSGKQOQAKYGGAPL 985
Qy 1013 BELQAKYBELFARKQISTGVPMDPSPVKIKQNSAKYFNGTILGRGAGYVLKQOATNTYFS 1072
Db 986 BELQAKYBELFARKQISTGVPMDPSPVKIQNSAKYFNGTILGRGAGYVLKQOATNTYFN 1045
Qy 1073 LVSDN---TFLPKSLVNPNGHTSSSVTGLVPDGKGYVYVYSTSGNOKNAFTSLGNWYVF 1129
Db 1046 -ISDNKEINFLPKTLN-----QDSQVGSYDGKGYVYVYSTSGYQAKNTFISEGDKWYVF 1099
Qy 1130 DNNGYMTVGAOSINGANYFFLSNGIQLRNAIYDNGNKLVSYYGNDGRYENGYYLF-GQQ 1188
Db 1100 DNNGYMTVGAOSINGVNYFFLSNGIQLRDAILKNEDGTAYYNGNDGRYENGYYQFMSGV 1159
Qy 1189 WRYFONGIMAVCLTRVHGAQVFDASGQAKQFITTADGKURYPDRSGNQOISNRFVN 1248
Db 1160 WRHFNNGEMSVGLTVIDGQVQYFDEMGYQAKGKFTVTADGKIRYFDKSGNMRYNRFTEN 1219
Qy 1249 SKGEWFLFDHNGVAVTGTVTENGRLYFKPKNVQAKGFEIRDANGYLYRYPDPSNGEVN 1308
Db 1220 EBGKWLVLGEDAAVTGSQTNGQHLFYFRANGVQVKGEFVTDHGHRI SYDNGSGDQIRN 1279
Qy 1309 RFRVNSKGEWFLFDHNGIAVTGARVYVNGHASTLSLMVFR 1347
Db 1280 RFRVNAQGWFFFDNNGVAVTGTARTING-----QLLYFR 1313

RESULT 14
AAU98032
ID AAU98032 standard; protein; 1475 AA.
XX
AC AAU98032;

XX 27-AUG-2002 (first entry)
DT
XX S. mutans glucosyltransferase GTFB mutant D567T.
DE
XX Glucosyltransferase; GTFB; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture; mutant; muten.
XX Streptococcus mutans.
OS Synthetic.
XX Key Location/Qualifiers
FH Misc-difference 567 /note= "Wild-type Asp substituted by Thr"
FT
XX US2002031826-A1.
PN
XX 14-MAR-2002.
PD
XX 19-DEC-2000; 2000US-00740274.
PF
XX 07-JUN-1995; 95US-00478704.
PR
XX 07-JUN-1995; 95US-00482711.
PR
XX 07-JUN-1995; 95US-00485243.
PR
XX 16-JAN-1998; 98US-00007999.
PR
XX 16-JAN-1998; 98US-00008172.
PR
XX 20-JAN-1998; 98US-00009620.
PR
XX 11-DEC-1998; 98US-00210361.
PA (NICH/) NICHOLS S E.
XX
XX Nichols SE;
PI
XX WPI; 2002-414332/44.
DR
XX Glucosyltransferase B or D protein useful for producing a glucan useful
PT as substitutes for and additions to modified starch and latexes in paper
PT manufacture, comprises mutations in specific positions.
XX
XX Claim 36; Page: 44pp; English.
CC The invention an isolated protein comprising a glucosyltransferase (GTF)
CC B polypeptide having changes at position from I48V, D457N, D567T,
CC K1014T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,
CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
CC GTF mutant, an isolated polynucleotide which encodes p1 or p2, or its
CC complementary polynucleotide, a ribonucleic acid sequence encoding the
CC GTF mutant, an expression cassette comprising the polynucleotide operably
CC linked to a promoter, a vector comprising the expression cassette, host
CC cell introduced with the vector, a transgenic plant comprising the
CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
CC coating composition comprising a glucan produced in a plant transformed
CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
CC thermoplastic molecule or their combinations or glucan and starch where
CC the glucan is produced in the amyloplast and/or vacuole or a maize line
CC deficient in starch biosynthesis, transformed with a gene encoding a
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
CC comprising the glucan (paper sizing/coating agent). The vector is useful
CC for producing a glucan in a plant. The method comprises transforming a
CC plant cell with the vector, growing the plant cell under plant growing
CC conditions to produce a regenerated plant and inducing expression of the
CC polynucleotide for a time sufficient to produce the glucan in the
CC regenerated plant, where the vector contains a transit sequence from
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
CC is produced in the amyloplast of potato or the vacuole of sugar beet.
CC Glucans are useful as substitutes for and additions to modified starch
CC and latexes in paper manufacture. Unlike prior art techniques, which
CC require input materials that produce chemical effluents, paper
CC manufacture utilising the glucan produced by GTF, which utilises

Search completed: February 11, 2006, 19:14:42
Job time : 141.863 secs

1000 400 200 00

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 19:39:30 ; Search time 30.1046 Seconds
(without alignments)
3776.130 Million cell updates/sec

Title: US-10-797-821-35
Perfect score: 7143
Sequence: 1 MEKKVRFKLKRVKRWVTS.....VKVSNMILIPMKFVIM 1375

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/prodata/1/iaa/5 COMB.pep: *
2: /cgn2_6/prodata/1/iaa/6 COMB.pep: *
3: /cgn2_6/prodata/1/iaa/H COMB.pep: *
4: /cgn2_6/prodata/1/iaa/PCTUS COMB.pep: *
5: /cgn2_6/prodata/1/iaa/RE COMB.pep: *
6: /cgn2_6/prodata/1/iaa/backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7143	100.0	1375	2	US-09-210-361-4
2	7143	100.0	1375	2	US-09-740-274-4
3	5285	74.0	1475	2	US-09-007-999-2
4	5285	74.0	1475	2	US-09-210-361-2
5	5285	74.0	1475	2	US-09-740-274-2
6	3591.5	50.3	1430	2	US-09-008-172-2
7	3591.5	50.3	1430	2	US-09-210-361-6
8	3591.5	50.3	1430	2	US-09-740-274-6
9	3216	45.0	1577	1	US-08-793-824-2
10	2501.5	35.0	2057	2	US-09-499-203-2
11	2347.5	32.9	1278	2	US-09-604-957-3
12	2347.5	32.9	1781	2	US-09-995-749A-2
13	1787	25.0	349	2	US-09-009-620-2
14	1718.5	24.1	545	2	US-09-604-957-4
15	1690.5	23.7	545	2	US-09-995-749A-10
16	1620.5	22.7	523	2	US-09-604-957-5
17	1616	22.6	522	2	US-09-995-749A-11
18	1445.5	20.2	535	2	US-09-604-957-7
19	1445.5	20.2	535	2	US-09-995-749A-13
20	1337	18.7	584	2	US-09-995-749A-12
21	1333	18.7	584	2	US-09-604-957-6
22	312.5	4.4	2710	1	US-08-480-604A-6
23	312.5	4.4	2710	1	US-08-405-496A-6
24	312.5	4.4	2710	2	US-08-915-136-6
25	312.5	4.4	2710	2	US-08-957-310-6
26	312.5	4.4	2710	2	US-10-011-366-6
27	312.5	4.4	2710	2	US-09-084-517-6

28	272	3.8	1742	2	US-09-386-962C-4	Sequence 4, Appli
29	272	3.8	1742	2	US-09-386-959-4	Sequence 4, Appli
30	269.5	3.8	811	1	US-08-480-604A-7	Sequence 7, Appli
31	269.5	3.8	811	1	US-08-405-496A-7	Sequence 7, Appli
32	269.5	3.8	811	2	US-08-915-136-7	Sequence 7, Appli
33	269.5	3.8	811	2	US-08-957-310-7	Sequence 7, Appli
34	269.5	3.8	811	2	US-10-011-366-7	Sequence 7, Appli
35	269.5	3.8	811	2	US-09-084-517-7	Sequence 7, Appli
36	269.5	3.8	812	1	US-08-480-604A-29	Sequence 29, Appli
37	269.5	3.8	812	2	US-08-915-136-29	Sequence 29, Appli
38	269.5	3.8	812	2	US-09-084-517-29	Sequence 29, Appli
39	268.5	3.8	2366	1	US-08-480-604A-10	Sequence 10, Appli
40	268.5	3.8	2366	1	US-08-405-496A-10	Sequence 10, Appli
41	268.5	3.8	2366	2	US-08-915-136-10	Sequence 10, Appli
42	268.5	3.8	2366	2	US-08-957-310-10	Sequence 10, Appli
43	268.5	3.8	2366	2	US-10-011-366-10	Sequence 10, Appli
44	268.5	3.8	2366	2	US-09-084-517-10	Sequence 10, Appli
45	266	3.7	1155	2	US-09-710-279-1780	Sequence 1780, Ap

ALIGNMENTS

RESULT 1
US-09-210-361-4
; Sequence 4, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; EARLIER FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; TYPE: PRT
; LENGTH: 1375
; ORGANISM: streptococcus mutans
US-09-210-361-4

Query Match	100.0%;	Score 7143;	DB 2;	Length 1375;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1375;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MEKKVRFKLKRVKRWVTSIASAVVTLTSLGSLVKADSTDDRQAVTESQASLVTSE	60	
Db	1	MEKKVRFKLKRVKRWVTSIASAVVTLTSLGSLVKADSTDDRQAVTESQASLVTSE	60	
QY	61	AAKETLTATDTSTATSQPTATVTDNVSTTNQSTNTTANTANFVVKPTTSEQAKTDN	120	
Db	61	AAKETLTATDTSTATSQPTATVTDNVSTTNQSTNTTANTANFVVKPTTSEQAKTDN	120	
QY	121	SDKIITTSKAVNRLTATGKVFANNNTAHPKVTVDKIVIPKIKGLKQPSSLSQDDIAA	180	
Db	121	SDKIITTSKAVNRLTATGKVFANNNTAHPKVTVDKIVIPKIKGLKQPSSLSQDDIAA	180	
QY	181	LGNVKNIRKNGKYYYKEDGTGLQKNYALNINGKTFFFDETGALSNNITLPSKKGNTNND	240	
Db	181	LGNVKNIRKNGKYYYKEDGTGLQKNYALNINGKTFFFDETGALSNNITLPSKKGNTNND	240	

```
QY 241 NTNSFAQYNQVSTVDVANFEHVDHYLTAEBSWYRPKYILKDGKTWTQSTOKDPRPLLMTWM 300
Db 241 NTNSFAQYNQVSTVDVANFEHVDHYLTAEBSWYRPKYILKDGKTWTQSTOKDPRPLLMTWM 300

QY 301 PQDETQRYVYNNMNAQLGHOYTNTATSPQLNLAAQTITQKIEEKITAENKTNWLRQTI 360
Db 301 PQDETQRYVYNNMNAQLGHOYTNTATSPQLNLAAQTITQKIEEKITAENKTNWLRQTI 360

QY 361 SAFVKTQSAWNSDSEKPFDDHLQKALLYNNKSLTSQANSNYRILNRPPTNQTKGKDR 420
Db 361 SAFVKTQSAWNSDSEKPFDDHLQKALLYNNKSLTSQANSNYRILNRPPTNQTKGKDR 420

QY 421 YTADRTIGGYEFLFLLANDVNSNPVQAEOLNHLFLMNFNGNIYANDPDANFOSIRVDAVD 480
Db 421 YTADRTIGGYEFLFLLANDVNSNPVQAEOLNHLFLMNFNGNIYANDPDANFOSIRVDAVD 480

QY 481 NVDADLLOLAGDYLAAGKGIHKNDKAANDHLSILEAWSYNDTPYLLHDDGDNMINMDNRLR 540
Db 481 NVDADLLOLAGDYLAAGKGIHKNDKAANDHLSILEAWSYNDTPYLLHDDGDNMINMDNRLR 540

QY 541 LSLYSLAKPLNORSQWNPILITNSLVNRTDDNAETAAPVPSYFIRAHDSVQDLIRNIIR 600
Db 541 LSLYSLAKPLNORSQWNPILITNSLVNRTDDNAETAAPVPSYFIRAHDSVQDLIRNIIR 600

QY 601 TEINPNVGYSTFTEEBIKAFELIYNKDLATEKKYTHYNTALSYALLTNKSVPRVYVG 660
Db 601 TEINPNVGYSTFTEEBIKAFELIYNKDLATEKKYTHYNTALSYALLTNKSVPRVYVG 660

QY 661 DMFTDDGQYMAHKTINYEAIETLLKARIKYVSGQAMRNQOVGNSBIITSVRYGKALKA 720
Db 661 DMFTDDGQYMAHKTINYEAIETLLKARIKYVSGQAMRNQOVGNSBIITSVRYGKALKA 720

QY 721 TDTGDRTRTSGVAVTEGNNPSRLKASDRVVVNMGAHKNQAYRPLLLTTTNGIKAYHS 780
Db 721 TDTGDRTRTSGVAVTEGNNPSRLKASDRVVVNMGAHKNQAYRPLLLTTTNGIKAYHS 780

QY 781 DQEAAGLVRYTNDRGELITAAADIKGYANPOVSGYLGWVPVGAADQDVRVAASAPST 840
Db 781 DQEAAGLVRYTNDRGELITAAADIKGYANPOVSGYLGWVPVGAADQDVRVAASAPST 840

QY 841 DGKSVHQNAAALSRVMFEGSFNFOAFATKKEEYTNVVIKNDVKFAEWGVTDFEMAPOVY 900
Db 841 DGKSVHQNAAALSRVMFEGSFNFOAFATKKEEYTNVVIKNDVKFAEWGVTDFEMAPOVY 900

QY 901 SSTDGSFLDSVIONGVAFTRDYDLGISKPNKYGTADDLVKAIKALHSKGIKVMADWVPDQ 960
Db 901 SSTDGSFLDSVIONGVAFTRDYDLGISKPNKYGTADDLVKAIKALHSKGIKVMADWVPDQ 960

QY 961 MYALPEKEVVYTRVDKYGTPVAGSQIKNTLYVVDGKSGKQOQAKYGGAPLEELQAKYP 1020
Db 961 MYALPEKEVVYTRVDKYGTPVAGSQIKNTLYVVDGKSGKQOQAKYGGAPLEELQAKYP 1020

QY 1021 ELFAKQISTGVPMDSVKIKOWSAKYFNGTNILGRGAGVLDQATNTYFSLVSDNTFL 1080
Db 1021 ELFAKQISTGVPMDSVKIKOWSAKYFNGTNILGRGAGVLDQATNTYFSLVSDNTFL 1080

QY 1081 PKSLVNPNGHGTSSVTLGVFDGKGYVYVYSTSGNQAKNAFISLGNWNYFDDNNGYMTGAQ 1140
Db 1081 PKSLVNPNGHGTSSVTLGVFDGKGYVYVYSTSGNQAKNAFISLGNWNYFDDNNGYMTGAQ 1140

QY 1141 SINGANYFLSNGIQLRNAIYDNGNKVLSYNGDGRYENGYYLFCQOQRYFQNGIMAYG 1200
Db 1141 SINGANYFLSNGIQLRNAIYDNGNKVLSYNGDGRYENGYYLFCQOQRYFQNGIMAYG 1200

QY 1201 LTRVHGAQVYFASGFOAKGQFITTADGKLRVFDKRGSGNQISNRFVRNSKGWFLFDHNG 1260
Db 1201 LTRVHGAQVYFASGFOAKGQFITTADGKLRVFDKRGSGNQISNRFVRNSKGWFLFDHNG 1260

QY 1261 VAVTGTVTNGQLRYFKPKNGVQAKGEFIRDANGYLRYDPSNGNEVNRNRFVRNSKGWFL 1320
Db 1261 VAVTGTVTNGQLRYFKPKNGVQAKGEFIRDANGYLRYDPSNGNEVNRNRFVRNSKGWFL 1320
```

```
QY 1321 FDHNGIANTGARVNVNGHASILSLMVFRLRESSLQSVKVVSNMTILIPMKFVIM 1375
Db 1321 FDHNGIANTGARVNVNGHASILSLMVFRLRESSLQSVKVVSNMTILIPMKFVIM 1375

RESULT 2
US-09-740-274-4
; Sequence 4, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PR1
; ORGANISM: streptococcus mutans
US-09-740-274-4
```

```
Query Match 100.0%; Score 7143; DB 2; Length 1375;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MEKKVRFKLKRVKKEWTVSIA5AVVTLTSLGSLVKADSTDDRQOAVTESQASLVTTSE 60
Db 1 MEKKVRFKLKRVKKEWTVSIA5AVVTLTSLGSLVKADSTDDRQOAVTESQASLVTTSE 60

QY 61 AAKETLTATDTSTATSATSQPTATVTDNVSTTNQSTNTTANTANFVVKPTTSEQAKTDN 120
Db 61 AAKETLTATDTSTATSATSQPTATVTDNVSTTNQSTNTTANTANFVVKPTTSEQAKTDN 120

QY 121 SKIIITTSKAVNRLTATGKGFVANNNTAHPKVTVDKIVIPKIGLKPSSLSODDIAA 180
Db 121 SKIIITTSKAVNRLTATGKGFVANNNTAHPKVTVDKIVIPKIGLKPSSLSODDIAA 180

QY 181 LGNVKNIRKVGKYYKEDGTQKNYALNINNGKTFPFDGTGALSNNTLPSKKGNITND 240
Db 181 LGNVKNIRKVGKYYKEDGTQKNYALNINNGKTFPFDGTGALSNNTLPSKKGNITND 240

QY 241 NTNSFAQYNQVSTVDVANFEHVDHYLTAEBSWYRPKYILKDGKTWTQSTOKDPRPLLMTWM 300
Db 241 NTNSFAQYNQVSTVDVANFEHVDHYLTAEBSWYRPKYILKDGKTWTQSTOKDPRPLLMTWM 300

QY 301 PQDETQRYVYNNMNAQLGHOYTNTATSPQLNLAAQTITQKIEEKITAENKTNWLRQTI 360
Db 301 PQDETQRYVYNNMNAQLGHOYTNTATSPQLNLAAQTITQKIEEKITAENKTNWLRQTI 360

QY 361 SAFVKTQSAWNSDSEKPFDDHLQKALLYNNKSLTSQANSNYRILNRPPTNQTKGKDR 420
Db 361 SAFVKTQSAWNSDSEKPFDDHLQKALLYNNKSLTSQANSNYRILNRPPTNQTKGKDR 420

QY 421 YTADRTIGGYEFLFLLANDVNSNPVQAEOLNHLFLMNFNGNIYANDPDANFOSIRVDAVD 480
Db 421 YTADRTIGGYEFLFLLANDVNSNPVQAEOLNHLFLMNFNGNIYANDPDANFOSIRVDAVD 480
```

Qy	481	NVDADLLQIAGDYLKAAKGIIHKNDKAANDHLSIILEANSYNDTPYLIHDDGDNMINDNRLR	540
Db	481	NVDADLLQIAGDYLKAAKGIIHKNDKAANDHLSIILEANSYNDTPYLIHDDGDNMINDNRLR	540
Qy	541	LSLLYSLAKPLNQRSGMNPILITNSLVNRTDDNAETAAPVPSYFIRAHDSVQDILIRNIIR	600
Db	541	LSLLYSLAKPLNQRSGMNPILITNSLVNRTDDNAETAAPVPSYFIRAHDSVQDILIRNIIR	600
Qy	601	TEINPNVVGYSFTTEEIKKAEIYNKOLLATTEKKYTHYNTALSYALLLTNKSVPVRYVG	660
Db	601	TEINPNVVGYSFTTEEIKKAEIYNKOLLATTEKKYTHYNTALSYALLLTNKSVPVRYVG	660
Qy	661	DMFTDDGQYMAHKTIYNEAIEITLLKARIKYYSGGQAMRNQOVGNSEIITSVRYKGKALKA	720
Db	661	DMFTDDGQYMAHKTIYNEAIEITLLKARIKYYSGGQAMRNQOVGNSEIITSVRYKGKALKA	720
Qy	721	TDTGDRTRTSGVAVIEGNPSSLRLKASDRVVNNGAAHNQAYRPLLLTTDNGIKAYHS	780
Db	721	TDTGDRTRTSGVAVIEGNPSSLRLKASDRVVNNGAAHNQAYRPLLLTTDNGIKAYHS	780
Qy	781	DOEAAGLVRYTNDRGELIFTAAIDIKGVANPOVSGVLGVWVPVGAADODVRVAASTAPST	840
Db	781	DOEAAGLVRYTNDRGELIFTAAIDIKGVANPOVSGVLGVWVPVGAADODVRVAASTAPST	840
Qy	841	DGKSVHQNAALDSRVMFEGFSNFQAFATKKEEYTNVVIKNVDKFAEWGVTDFEMAPOYV	900
Db	841	DGKSVHQNAALDSRVMFEGFSNFQAFATKKEEYTNVVIKNVDKFAEWGVTDFEMAPOYV	900
Qy	901	SSTDGSLDSVIONGYAFTDRYDILIGSKPNKYGTADDLVKAIKALHSGIKVMADWPYDQ	960
Db	901	SSTDGSLDSVIONGYAFTDRYDILIGSKPNKYGTADDLVKAIKALHSGIKVMADWPYDQ	960
Qy	961	MYALPEKEVVTATVDKYGTPVAGSQIKNTLIYVDGSSGKDQQAQYGGAFLEELQAKYP	1020
Db	961	MYALPEKEVVTATVDKYGTPVAGSQIKNTLIYVDGSSGKDQQAQYGGAFLEELQAKYP	1020
Qy	1021	ELFARKQISTGVPMDPSVKIKQWSAKYPNGTNILGRGAGYVYLKDOATNTYPSLSVSDNTFL	1080
Db	1021	ELFARKQISTGVPMDPSVKIKQWSAKYPNGTNILGRGAGYVYLKDOATNTYPSLSVSDNTFL	1080
Qy	1081	PKSLVNPENHGTSSSVTGLVPDGKGVVYYSYSGNOAKNAFISLGNWYVYFDNNGYVMTGAQ	1140
Db	1081	PKSLVNPENHGTSSSVTGLVPDGKGVVYYSYSGNOAKNAFISLGNWYVYFDNNGYVMTGAQ	1140
Qy	1141	SINGANYFYLSNGIQLRNAIYDNGNKVLSYDNGDRRYENGYYILFGQWRYPFQNGIMAVG	1200
Db	1141	SINGANYFYLSNGIQLRNAIYDNGNKVLSYDNGDRRYENGYYILFGQWRYPFQNGIMAVG	1200
Qy	1201	LTRVHGAVQYFDASGFOAKGQFITTAGDKLRYFDRDSGNQISNRPFRVNSKGWFLPDHNG	1260
Db	1201	LTRVHGAVQYFDASGFOAKGQFITTAGDKLRYFDRDSGNQISNRPFRVNSKGWFLPDHNG	1260
Qy	1261	VAVTGTVTFNQORLYFXPNPGVQAKGEFIRDANGVLYRYDNPNSGNEVNRNRPVRNSKGWFL	1320
Db	1261	VAVTGTVTFNQORLYFXPNPGVQAKGEFIRDANGVLYRYDNPNSGNEVNRNRPVRNSKGWFL	1320
Qy	1321	FDHNGIATVGARVNVNGHASILSLMVPRLRBSLSQSVKVSNTMTILIPEMKPFVIM	1375
Db	1321	FDHNGIATVGARVNVNGHASILSLMVPRLRBSLSQSVKVSNTMTILIPEMKPFVIM	1375

RESULT 3

	US-09-007-999-2	
Reason 3	; Sequence 2, Application US/09007999	
	; Patent No. 6087559	
	; GENERAL INFORMATION:	
	; APPLICANT: Nichols, Scott E.	
	; TITLE OF INVENTION: Substitutes for Modified Starch and	
	; FILE OF INVENTION: Latexes in Paper Manufacture	
	; FILE REFERENCE: 0356D	
	; CURRENT APPLICATION NUMBER: US/09/007,999	
	; CURRENT FILING DATE: 1998-01-16	
	; EARLIER APPLICATION NUMBER: 08/478,704	


```
Db 1100 DNNGYMTGAQINGVNYFSLNGLQLRDLAKNEDGTAYAYGNDGRRYENGYYQFMGSV 1159
Qy 1189 WRYPQNGIMAVGLTRVHGAQVYFDASGFQAKGQFITTADGKRLYFRDRDSGNQISNRFVRN 1248
Db 1160 WRHFNNGEMSVGLTVIDGQVQYFDEMGYQAKGKEVTTADGKIRYFDKQSGNMYRNFIE 1219
Qy 1249 SKGEMFLPDHNGAVTGTTFNGQRLYFKPKNVQAKGEFIRDANGYLYRYDPPNSGNEVRN 1308
Db 1220 EBGKWLVLGSDGAATVTSQINGQHLFRANGVQVKGFEVTDHHRISYYDNGSGDQIRN 1279
Qy 1309 RFVRNSKGEWFLPDHNGIATVGTARVNGHASILSLMVF 1347
Db 1280 RFVRNAQGFYFDNNGYAVTGARTING-----QLLYFR 1313

RESULT 5
US-09-740-274-2
; Sequence 2, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-740-274-2

Query Match 74.0%; Score 5285; DB 2; Length 1475;
Best Local Similarity 75.9%; Pred. No. 0;
Matches 1031; Conservative 101; Mismatches 169; Indels 58; Gaps 14;

Qy 1 MEKKVRFKLRKVKRWVTVSIAASAVTLTSLGSLVKADSTDDROQAVTESQASLVTSE 60
Db 1 MDKKVIRKLRKVKRWVTVSIAASAVTLTSLGSLVKADSNESKQISNDSNTSVVTANE 60

Qy 61 AAKETLTATDTSTATSATSOPTA--TWTDNVSTFNQSTNTNTANTANFVVKPTTTSBQAKT 118
Db 61 -----ESNVITEATSKQEAASQTNHTVTTSSSSSVVNPKEVSNPYTVGETA-- 109

Qy 119 DNSDKLIITTSKAVNRLTATGKFVPANNHTAHPKVTVDKIVPKTKGLKQPSLSQDDI 178
Db 110 SNGEKLQNTTVDKTSSEA-----AANNISKOTTEAD-----TDVIDDSEN 149

Qy 179 AA-----LGNVKIRKNGKVVYKEDGTQLQKVALNKGKTFPFDGTGALSNTLPSKKG 234
Db 150 AANLQILEKLPNKVIEDGKYIYDNNKVRNFTLIADGKILHFDGTGAYTDTSDITVVK 209

Qy 235 NITNNDNTNSFAQYNQVSTDVANFEHVDHYLTAEWSYRPKYTLKDGKTWTOSTEKDFRP 294
Db 210 DIVTT--RSNLYKYNQVDRSAQSFEHVDHYLTAEWSYRPKYTLKDGKTWTOSTEKDFRP 268

Qy 295 LMTWMPDQETQRQYVNYMNAQLGIHOTYNTATSPQLNLAAQTIQTKIEBKITAENKTN 354
|||||
```

RESULT 6
US-09-008-172-2

```
Db 269 LMTWMPDQETQRQYVNYMNAQLGIKNTYDDTNSQLQNLNIAAAATQAKIEAKITTLKNVT 328
Qy 355 WLQRTISAFVKTSQAWNSDEKFPDDHLQKGLLYSNNSKLTSQANSNYRILNRTPTNQT 414
Db 329 WLQRTISAFVKTSQAWNSDEKFPDDHLQNGAVLYDNEGKLTYPYANSNYRILNRTPTNQT 388
Qy 415 GKXDPRTADRTTGGYEFFLLANDVDSNPNVQAEQLNLWHLFNMFGNIYANDPDANFDSI 474
Db 389 GKXDPRTADRTTGGYEFFLLANDVDSNPNVQAEQLNLWHLFNMFGNIYANDPDANFDSI 448
Qy 475 RVDAVNVDADLLQIAGDYLKAAKGIHKNDKAANDHLSILEANSYNDTPYLHDDGDMNIN 534
Db 449 RVDAVNVDADLLQIAGDYLKAAKGIHKNDKAANDHLSILEANSYNDTPYLHDDGDMNIN 508
Qy 535 MDNRLRLSLYLAKPLNORSNMNPLITNSLVNRTDDNAETAAPSVPSYFTRAHDSVQDL 594
Db 509 MDNKLRLSLFLSLAKPLNQRSNMNPLITNSLVNRTDDNAETAAPSVPSYFTRAHDSVQDL 568
Qy 595 IRNIIRTEINPNVVGYSFTTEEIKKAFIYNKOLLATEKKYTHYNTALSVALLLTNKSSV 654
Db 569 IADIIKAEINPNVVGYSFTMEEIKKAFIYNKOLLATEKKYTHYNTALSVALLLTNKSSV 628
Qy 655 PRVYVGDMFTDDGQYMAHKTINYEAJETLLKARIKYVSGGQAMRNQOVGNSEIITSRYG 714
Db 629 PRVYVGDMFTDDGQYMAHKTINYEAJETLLKARIKYVSGGQAMRNQOVGNSEIITSRYG 688
Qy 715 KGALKATDGTDRITRTSGVAVIEGNPNPSRLKASDRVVNMGAHKNQAYRPLLLTTDNG 774
Db 689 KGALKATDGTDRITRTSGVAVIEGNPNPSRLKASDRVVNMGAHKNQAYRPLLLTTDNG 748
Qy 775 IKAYHSDQEAAGLVRYTNDRGELIFTAADI KGYANPQVSGYLGWVWPVGAADQ--DVRV 832
Db 749 IKAYHSDQEAAGLVRYTNDRGELIFTAADI KGYANPQVSGYLGWVWPVGAALIKMFALRL 808
Qy 833 AASTAPSTDGKSVHQAALDSDRVMEGFSNFQAFATKBEYTNVVIKQNVDFKAEGWVTD 892
Db 809 A---RPHQQWASVHQNAALDSDRVMEGFSNFQAFATKBEYTNVVIKQNVDFKAEGWVTD 865
Qy 893 FEMAPQVSSDTGSLDQSVIQNGYAFTRDYDLGISKPNKYGTADDLKVAKALHSHKGIKV 952
Db 866 FEMAPQVSSDTGSLDQSVIQNGYAFTRDYDLGISKPNKYGTADDLKVAKALHSHKGIKV 925
Qy 953 MADWVPDQMYALPEKEVVTATRVDKYTPVAGSQIKNTLYVVDGSKSGKQQAQYGGAF 1012
Db 926 MADWVPDQMYALPEKEVVTATRVDKYTPVAGSQIKNTLYVVDGSKSGKQQAQYGGAF 985
Qy 1013 EELQAKYPFLPARKQISTGVPMDDPSVKIKQWAKYFNGTNILGRGAGYVLKQATNTYFS 1072
Db 986 EELQAKYPFLPARKQISTGVPMDDPSVKIKQWAKYFNGTNILGRGAGYVLKQATNTYFN 1045
Qy 1073 LVSDN---TFLPKSLVNPNHGTSSTSVTLGLVFDGKGVVYVYSTGNOAKNAFISLGNWYF 1129
Db 1046 -ISDNKEINFLPKTLN-----QDSQVGSYDGKGVVYVYSTGQAKNTFISEGDKWYF 1099
Qy 1130 DNNGYMTGAQINGANYFYFLSNGQLRNLAIYDNGKNVLSYNGDGRYENGYYLF-GQQ 1188
Db 1100 DNNGYMTGAQINGANYFYFLSNGQLRDLAILKNEDGTAYAYGNDGRRYENGYYQFMGSV 1159
Qy 1189 WRYPQNGIMAVGLTRVHGAQVYFDASGFQAKGQFITTADGKRLYFRDRDSGNQISNRFVRN 1248
Db 1160 WRHFNNGEMSVGLTVIDGQVQYFDEMGYQAKGKEVTTADGKIRYFDKQSGNMYRNFIE 1219
Qy 1249 SKGEMFLPDHNGAVTGTTFNGQRLYFKPKNVQAKGEFIRDANGYLYRYDPPNSGNEVRN 1308
Db 1220 EBGKWLVLGSDGAATVTSQINGQHLFRANGVQVKGFEVTDHHRISYYDNGSGDQIRN 1279
Qy 1309 RFVRNSKGEWFLPDHNGIATVGTARVNGHASILSLMVF 1347
Db 1280 RFVRNAQGFYFDNNGYAVTGARTING-----QLLYFR 1313
```

```
; Sequence 2, Application US/09008172
; Patent No. 6127602
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 03580
; CURRENT APPLICATION NUMBER: US/09/008,172
; CURRENT FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: Streptococcus mutans
; US-09-008-172-2

Query Match          50.3%; Score 3591.5; DB 2; Length 1430;
Best Local Similarity 51.9%; Pred. No. 9.6e-254;
Matches 721; Conservative 215; Mismatches 350; Indels 103; Gaps 22;

QY 1 MEKKVRFKLRKVRKVVYSIASAVVTL-TLSGSLVKADS-----TDDROA 47
DB 1 METKRYKMKVKKVVTVAVASGLITLTTILGSSVSAETEQOTSDKVVTKQSEDDKA 60
QY 48 VTESQASLVTTSEA-AKETLTATDTSTATSQPTATVTDNVSTNQ-----TNTTAN 101
DB 61 SESSOTDAPKTAQTEQTAQSOQANVADTSTSIKETPSQITTOANSDDKTIVNTKSE 120
QY 102 TANFVVVKPTTSEQAKTNSDKLIITTSKAVNRLTATGKVPANNWTAHPKTVTKIVPK 161
DB 121 EAQTSERTKQSEEAQTASSQALTOAKA--BLT----- 152
QY 162 PKIGKLPQSSLSQD-----DIAALGNVKNIRKNGKYVYKEDGTLQKNVAINLNGKTF 216
DB 153 -----KQRTAAQAKENPVDLAAIPNVK---QIDGKYIIGSDGPKPKNFALTNNKVL 203
QY 217 PFDE-TGALSNNTLPSSKGNIT--NNDNTNSFAQYQNVYSTDVANPEHVDHYLTABS 273
DB 204 YFDKNTGALTDTTSQYQKQGLTKLND-----YTPHQIIVNFENTSLETIDNVVTADSWY 259
QY 274 PKYILKDGKTWTQSTKEDFRPLMTWPPQETQORQVNTVNAQ-LGIHOTYNTATSPLOJ 332
DB 260 PKDILKNGKTWTASSESDRLPLMSWPPDKQTQIAYLNTYMQOGLGTGENYTADSSQESL 319
QY 333 NLAAQTIOTKIEKTAENKTNWLRQTI SAFVKTSAMNSDSEKPPD-----DHLQKALL 388
DB 320 NLAAQTIVQVKEIKISQTOQWLRIINSFVKTPQNNWSQTESDTSAGEKDHLOGGALL 379
QY 389 YSNNSKLTQANSNYRILNRTPTNQTGKDPRTADRTTIGGYEFFLLANDVDNSNPVQAE 448
DB 380 YSNSDK-TAYANSDYLLNRTPTSTQTK--PKYFEDNSSGGYDFLLANDIDNSNPVQAE 436
QY 449 QLNWHLFLNPNIGNIYVANDPANDFSDIRVDVNDADILQIAGDYLLKAAKHKNDKAA 508
DB 437 QLNWHLFLNPNYISVANDPEANPDGVRVDVNDVNDADLQIASDYLLKAHYGVKSEKNAI 496
QY 509 DHLSTLEANSYNDPTVYLHDGDNMIMNDRLLSLILYSLAKPLNQ-----RSGMNP 560
DB 497 NHLSTLEANSNDNDPQYKNTKGALPDKLRLSLYALTRPLEKDAKNEIRSGLEPV 556
QY 561 ITNSLNRDNDNAETAAPVSFIRAHDSVQDLRIIRIETINPNVVGYSFTTEIKKA 620
DB 557 ITNSLNRNRAEGKNSEMANIYFIRAHDSVQTVIAKIIKAQINPKPTDGLTFTLDELKQA 616
QY 621 FEIYNKDLATEKKYTHYNTALSYALLLTNKSQSVPRVYTGDMFTDDGQMAKTIYNAI 680
DB 617 FRIYNEDMRQAKKKYQSNIPYAYALMLSKNDISITRLYYGDMYSDDGQYMATKSPYDAI 676
QY 681 ETLKARIKIVSGGQAMRNQOVGNSE-----IITSVRYGKGAKATDGTDRTRTS 731
```

```
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-210-361-6

Query Match      50.3%; Score 3591.5; DB 2; Length 1430;
Best Local Similarity 51.9%; Pred. No. 9.6e-254;
Matches 721; Conservative 215; Mismatches 350; Indels 103; Gaps 22;

QY 1 MEKKVRFLKRVKRWVTSIGIAVVTL-TSLGSLVKADS-----TDDRQQA 47
DQ 1 METKRYKMEKVKHGWTVAVASGLITLTTTLGSSVSAETEQTSDKVTQKSEDDKAA 60
QY 48 VTSQASLVTTSEA-AKELTATDTSTATSOTPAITVDNVTNNQS-----TNTAN 101
DQ 61 SSSQTDAPKTKQAQTEBQAOQANVADTSITTKETPSQNTTQANSDDKTWTNKSE 120
QY 102 TANFVVKPTTTSQAQKTDNSDKIITTSKAVNRLTATGKFPANNNTAHPKTVTDKIVPIK 161
DQ 121 EAQTSBERTQSBQAQTTASSQAQTAQA--ELT----- 152
QY 162 PKIGKIKQPSLSQD-----DIAALGNVKNIRKVGKYYKYKEDGTLOKNYALNINNGKTF 216
DQ 153 -----KQRTAQAKENKPNVLAAPNVK---QIDGKYIIGSDGQPKKNFALTNNKVL 203
QY 217 FPDE-TGALSNNLTPLSKGNIT--NNDNTNSFAQYQVYSTDVANFEHVDHYLTAESWYR 273
DQ 204 YFDKNTGALTDTSQYQFKQGLTKLND---YTPHNQIVNFENTSLETIDNYVTADSWYR 259
QY 274 PKYILKDGKWTOSTEKDFPILMTWPDQETQROQVYNNYNAQ-LGIHQTYNTATSPQLQ 332
DQ 260 PKDILKNGKWTWATSSSDRLPPLMSWPDQKQIAYLYNNMQOGLGTGENYTTADSSQESL 319
QY 333 NLAAQTIQTKIEBKITAEKNTNMLROTISAFVKTOSAWNSDSKPPD----DHLQKALL 368
DQ 320 NLAAQTVQVKIEKISQTOQTQWLRLIINSFVKIQPNWNSQTSBDSAGESKHQLOGALL 379
QY 389 YSNNSKLTQANSRYILNRTPTNQTGCKDPRYTADRTTGGYFELLANDVDSNPVVQAE 448
DQ 380 YSNSDK-TAYANSDYRLNRTPTTSQTK--PKYFEDNSSGGYDFELLANDIDNSNPVVQAE 436
QY 449 QLNWLHFLNFGNIYANDPANFDSIRVADVNDADLQIAGDYILKAAKGHIHNDKAA 508
DQ 437 QLNWLHLYNMGYSIVANDPEANPDGVRVADVNDVNDADLQIAGDYILKAHYGVDSKNAI 496
QY 509 DHLIILEANSYNDTPYLHDDGDNMNMNDRRLSLYSLAKPLNQ-----RSCMNPL 560
DQ 497 NHLIILEASDNDPQNKQTKGQLPDKRLRLSLYALTLPLEKDAASNKEIRSGLEPV 556
QY 561 ITNSLVNRTDDNAETAAPVSYSPIRAHDSVQDLIRNIIRTEINPNVVGYSFTTEIRKKA 620
DQ 557 ITNSLVNRSAGKNSERMANYIFIRAHDSVQTVIAKIIKAQINPKTDGLTFLDELKQA 616
QY 621 FEIYNKDLATEKKYTHYNTALSYALLTNKSSVPVYGDMPDQGMMAHKTINYEAI 680
DQ 617 FKTYNEDMRQAKKYYTQSNIPYAYALMLSNKDSITRLYGYDMSDDGQYMATKSPYDAI 676
QY 681 ETLLKARIKYVSGGAWRNQVGNSE-----IITSVRYCKGALKATDTGDRTRTS 731
DQ 677 DTLKARIKYAGGQDMKITIYEGDQSHMDWDVTGVLTSVRGTGANEATDQGEATKQ 736
QY 732 GVAVIEGNNPSRLKASDRVVMNGAAHKNQAYRPLLLTTDNGIKAYHSDQEAAGLVRVT 791
DQ 737 GNAVITSNNPSRLKLNQNDKVINVMNGAAHKNQAYRPLLLTTKGLTSYTDAAAKSLYRKT 796
QY 792 NDRGELIFTADIKGYANPQVSGYLGWVPVGAADQDVRVAASTAPSTDGKSVHQAAL 851
DQ 797 NDKGELVFPASDIQGYLNFQVSGYLAVMVPVGASDNDQVRVAASNKANATGQVYESSAL 856
QY 852 DSRVMEGESNFQAFAFKKEETNNVIAKNVQKFAEWGTYTDFEMAPQYVSSDTGSLDSV 911
DQ 857 DSQLIYEGFSNFQDFVTTKDSYTNKKIAQNVLFKSWGVTSEFEMAPQYVSSDGSFLDSI 916

RESULT 8
US-09-740-274-6
; Sequence 6, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-6

Query Match      50.3%; Score 3591.5; DB 2; Length 1430;
Best Local Similarity 51.9%; Pred. No. 9.6e-254;
Matches 721; Conservative 215; Mismatches 350; Indels 103; Gaps 22;

QY 1 MEKKVRFLKRVKRWVTSIGIAVVTL-TSLGSLVKADS-----TDDRQQA 47
DQ 1 METKRYKMEKVKHGWTVAVASGLITLTTTLGSSVSAETEQTSDKVTQKSEDDKAA 60
QY 48 VTSQASLVTTSEA-AKELTATDTSTATSOTPAITVDNVTNNQS-----TNTAN 101
DQ 61 SSSQTDAPKTKQAQTEBQAOQANVADTSITTKETPSQNTTQANSDDKTWTNKSE 120
QY 102 TANFVVKPTTTSQAQKTDNSDKIITTSKAVNRLTATGKFPANNNTAHPKTVTDKIVPIK 161
DQ 121 EAQTSBERTQSBQAQTTASSQAQTAQA--ELT----- 152
QY 162 PKIGKIKQPSLSQD-----DIAALGNVKNIRKVGKYYKYKEDGTLOKNYALNINNGKTF 216
DQ 153 -----KQRTAQAKENKPNVLAAPNVK---QIDGKYIIGSDGQPKKNFALTNNKVL 203
QY 217 FPDE-TGALSNNLTPLSKGNIT--NNDNTNSFAQYQVYSTDVANFEHVDHYLTAESWYR 273
DQ 204 YFDKNTGALTDTSQYQFKQGLTKLND---YTPHNQIVNFENTSLETIDNYVTADSWYR 259
QY 274 PKYILKDGKWTOSTEKDFPILMTWPDQETQROQVYNNYNAQ-LGIHQTYNTATSPQLQ 332
DQ 260 PKDILKNGKWTWATSSSDRLPPLMSWPDQKQIAYLYNNMQOGLGTGENYTTADSSQESL 319
QY 333 NLAAQTIQTKIEBKITAEKNTNMLROTISAFVKTOSAWNSDSKPPD----DHLQKALL 368
DQ 320 NLAAQTVQVKIEKISQTOQTQWLRLIINSFVKIQPNWNSQTSBDSAGESKHQLOGALL 379
QY 389 YSNNSKLTQANSRYILNRTPTNQTGCKDPRYTADRTTGGYFELLANDVDSNPVVQAE 448
DQ 380 YSNSDK-TAYANSDYRLNRTPTTSQTK--PKYFEDNSSGGYDFELLANDIDNSNPVVQAE 436
QY 449 QLNWLHFLNFGNIYANDPANFDSIRVADVNDADLQIAGDYILKAAKGHIHNDKAA 508
DQ 437 QLNWLHLYNMGYSIVANDPEANPDGVRVADVNDVNDADLQIAGDYILKAHYGVDSKNAI 496
QY 509 DHLIILEANSYNDTPYLHDDGDNMNMNDRRLSLYSLAKPLNQ-----RSCMNPL 560
DQ 497 NHLIILEASDNDPQNKQTKGQLPDKRLRLSLYALTLPLEKDAASNKEIRSGLEPV 556
QY 561 ITNSLVNRTDDNAETAAPVSYSPIRAHDSVQDLIRNIIRTEINPNVVGYSFTTEIRKKA 620
DQ 557 ITNSLVNRSAGKNSERMANYIFIRAHDSVQTVIAKIIKAQINPKTDGLTFLDELKQA 616
QY 621 FEIYNKDLATEKKYTHYNTALSYALLTNKSSVPVYGDMPDQGMMAHKTINYEAI 680
DQ 617 FKTYNEDMRQAKKYYTQSNIPYAYALMLSNKDSITRLYGYDMSDDGQYMATKSPYDAI 676
QY 681 ETLLKARIKYVSGGAWRNQVGNSE-----IITSVRYCKGALKATDTGDRTRTS 731
DQ 677 DTLKARIKYAGGQDMKITIYEGDQSHMDWDVTGVLTSVRGTGANEATDQGEATKQ 736
QY 732 GVAVIEGNNPSRLKASDRVVMNGAAHKNQAYRPLLLTTDNGIKAYHSDQEAAGLVRVT 791
DQ 737 GNAVITSNNPSRLKLNQNDKVINVMNGAAHKNQAYRPLLLTTKGLTSYTDAAAKSLYRKT 796
QY 792 NDRGELIFTADIKGYANPQVSGYLGWVPVGAADQDVRVAASTAPSTDGKSVHQAAL 851
DQ 797 NDKGELVFPASDIQGYLNFQVSGYLAVMVPVGASDNDQVRVAASNKANATGQVYESSAL 856
QY 852 DSRVMEGESNFQAFAFKKEETNNVIAKNVQKFAEWGTYTDFEMAPQYVSSDTGSLDSV 911
DQ 857 DSQLIYEGFSNFQDFVTTKDSYTNKKIAQNVLFKSWGVTSEFEMAPQYVSSDGSFLDSI 916

912 IQNGYAFPTDRYDLGISKPKNKYGTADDLVKAIALKHSKGIKVMADWPDQMYALPEKEVVT 971
917 IQNGYAFEDRYDLAMSKNNKYGSQODMINAVKALHKSIGIQVIADWVFDQIYNLPGKEVVT 976
972 ATRVDKYGTFVAGSOIKNTLYVVDGSKSDQQAQYGGAFLELOAKYKPELFAKQISTG 1031
977 ATRVNDYGEYKDSKSEIKNTLYAAANTKSNKDYQAKYGGAFLELAAYKPSIFNRTQISNG 1036
1032 VPMDDPSVKIKQWSAKYPNGTILGRGAGYVYLKDOATNTYFSLVSDNTFLPKSLVNPNGHT 1091
1037 KKIDPEKITAWAKAFNGTILGRGVYVYLNKNASDKYFELKGNQTYLTPKQMTN----- 1091
1092 SSSVTGLVDFGKGYVYVYSTSGNOAKNAFI-SLGNMNYIFDNGYMTVTAQSIINGANYFYL 1150
1092 KEASTGVNDGNGMTFYSTSGYQAKNSFVQDAKGNMNYIFDNGHMYVYGLQQLNGEVYFL 1151
1151 SNGIOLRNALYDNGNKVLSYNGDGRYENGYLF--GQWRYFO-NGIMAVGLTRVHGA 1207
1152 SNGVQURESFLENADGSKNYFGLHGNRYSGNYSPDNDSKWRYFDSAGVMVAVGLKTINGN 1211
1208 VOYFDASGFOAQGFITTDGKLYRFDROSGNQISNRFVRNSKGEWFLPDHNGVAVTGTV 1267
1212 TQYFDQGGYQVKGAWITGSKKRYFDDGSGNMAVRPANDKNGDMTYLNSDGLALVGVQ 1271
1268 TFGQRLYFKPQGVQAKGEFIRDANGLYRYDPNSGNEVRNFRVRSKGEWFLFDHNGIA 1327
1272 TINGKTYFQDQKQIKGKIITD-NGKLYFLANSBELARNIFATDSQNNWYFSGDGA 1330
1328 VTGARVNG 1336
1331 VTGSQTIA 1339
```



```
QY 50 ESQASLVTTTS-----EAAKETLTATDTST-----ATSATSO-----PTAT-- 84
Db 59 DGTASLVTTTTTTEQASQAASAVATASVSHSETSFQAATSASQBATQAQATSPVASOE 118
QY 85 --VTDNVSTTNOSTNTT-----ANTANFVVKPT-----TTSEQAK-----TDNSDKIITTS 128
Db 119 VAVSSQTQSOGGETTTEQVSOQOTSTQVAGTSAQTSVTEQARPRVLTWAAPIATR 178
QY 129 KA-----VNRLTATGKVFVANNNTAHPKTPTDKIVPIKPKI----- 164
Db 179 AADSTIRINANRNTIITASGTTTPNVTIITGPNTB-KENVVTVTSPNGTRPNVTIIVTQPN 237
QY 165 --GKLKQPSLSODDIAALGN-----VNIRKVNKGYYYKEDGTLOKNYALNING 213
Db 238 QNKPVPQSPQSPQPNKVPQNPQPSLDYKFPVSNLKTIDGKYV-ENGVVVKQNAIELDG 296
QY 214 KTFPFDETCALSNNTLPSKKGN-ITNNDNTNFAQVNOVYSTDVANFEHVDHYLTAESWY 272
Db 297 RLYYFDETGAMVDQSKPLRYADAIPNN---SIYAVNQAYDTSSKSFELHDLNFLTADSWY 353
QY 273 RPKYILKDGKTWTQSTEKDFRPLMTWMPDQBTQORQYVNMNAOLGIHQTYNTATSPLOL 332
Db 354 RPKQILKDGKNWTASTEKDFRPLMTWMPDKVTQVNYLNMYSQQGFGNKTYITDMMSYDL 413
QY 333 NLAQOTIQTKIEBKITAEXNTWLTQTSIAFVKTSQAMNSDSEKFP---DDHLQKALLY 389
Db 414 AAAAETVQRIEIEREGNCTWLRQMSDFIKTQPGWNSSESDNLLVGKDHLOGGALTFF 473
QY 390 SNNSKLTQSAQNYRILNARTPTNQTGKDPRTADRTIGGYEFLLANDVNSNPVVQABQ 449
Db 474 LNNs-ATSHANDFRMLNKPNTNQGTTR--KYHIDRSNGGYELLANDIDNSNPVQABQ 530
QY 450 LNLWHLFMNFGNIYADNPANFDSIRVDVNDVADLLOIAGDYLAKAAGIHKNDKAAND 509
Db 531 LNLWHLVIMNIGSILGNDPSANFDGVRIDAVDNVDADLLOIAGDYFKEKYRVADNEANAIA 590
QY 510 HLSILEAWSYNDTPYLHDDGDMNIMNDRLSLUSLYSLAKPLNQSGMNPITNSLVNRT 569
Db 591 HLSILEAWSYNDHQNKDTKGALSDINPLRETLTTFLURKSNYRGSLERVITNSLNNRS 650
QY 570 DNNAETAAPVPSYFIRANDSEVODLIRNIIRTEINPNVGVYFTTTEEIKKAFIYNKOLL 629
Db 651 SEQKHTPRDANFYFRAHDSVQAVLANIISKOINPKTDGFTTMDLQKAFIYNADIA 710
QY 630 ATEKKYTHYNTALSALLTNKSSPRVYVYGMFTDDGOYMAHKTINTYEAETILKARIK 689
Db 711 KADKKYTOYNIPAAVATMLNKDSITRYVYGLDFTDDGOYMAEKSPYNAIDALLARIK 770
QY 690 YVSGQAMRNQOVGNSEIITSVRYGKALKATDTGDRTRTSQVAVIEGNPNPSLRLKASD 749
Db 771 YVAGQDMKVTKLNGYEIMSSVRYGKGAEEANQLGTAETRNQGLMLVLTANRPDMKLGAND 830
QY 750 RVVNMGAHKNQOARPEILLITTDNGIKAYHSDQEA-AGLVRYTNDRGBELIFTAADIKGYA 808
Db 831 RLWNMGAHKNQOARPEILLKSTGLATYLDKSDVPAGLVRYTDNQGMLTFTTADDIAGHS 890
QY 809 NPQVSGYLGWVVPVGAADQDVRVAASTAPSTDGKSVHQ--NAALDSRVNMFEGFSNFOAFA 867
Db 891 TVEVSGYLAWVPVGAENQDARTKASS--TKKGQVFESSAALDOSQVIEGFSNFQDFV 948
QY 868 TKKEBYTNVIAKNVDKFAEMGVTDFFEMAPQVYSSTDGSLDSVIONGYAFTRDYDLGIS 927
Db 949 KTPSQYTNRVIAQNAKLFKEWIGITSFEPAPQVYSSQDGTFLDLSIENGVAFEDRYDIAMS 1008
QY 928 KPNKYGTADDLVKAIKALHSKGIKWADWPDPOMYALPEKEVVTATRVDKYGTVPVAGSQI 987
Db 1009 KNNKYGSLKOLMDLALHAEGISAIAADWPVQIYNLPKGEVVTYASRTNSYGTTPRPAEI 1068
QY 988 KNTLVVVDGSKGQDQQAQYGAFLLELOAKYPELFAKQISTGVPMDPSPVKIKOWSAKY 1047
Db 1069 YNSLAAKTRTFGNDFOQYGAFLJELKAKYPAIFERVQISNGRKLTTNEKITSWSAKY 1128
QY 1048 FNGTNILGRGAGYVLKQOATNTYFSLVSDNTFLPKSLVNPNHGTSSTSVTGLVFD--GKGY 1105
```

```
Db 1129 FNGSNIQGTGARYVLQDNATNOYFSVKAGOTFLPKQM-----TEITSGSPRRVGDDV 1180
QY 1106 VYVYSTGNOAKNAFISLIG--NMWYFDDNGYMTVGAQSIINGANYFPLSNGIQLRNALYDNG 1164
Db 1181 QYLSIGGYLAKNTFTQVGNQWYFDDKNGWMTGEVIDGKYFFLDNGLQLHLVLRQGS 1240
QY 1165 NKVLSYNGDGRYRYENGYYLFG---QOMRYFO-NGIMAVGLTRVHGAVOYFD-ASGFOAK 1219
Db 1241 DGHVYVYDPKGVQAFNGFYDFAGPRQDVRYPDNGQMYRGLHDMYGTTFYFDEKTCGIAK 1300
QY 1220 GOFITTADCKLRYFORDSGNOISNRPVRNSKGE-WFLPDHNGVAVTGTTFNGQRLYFKP 1278
Db 1301 DKFIRFADGRTRYFTPDTCNLAVNRFAQNPENKAWYILDSNGYAVTGLQTINGKQYFFDN 1360
QY 1279 NGVQAKGFFIRDANGVLYRY-DPNSGNEVRNRPVRNSKGEWFLPDHNGTAVTGARVVNGH 1337
Db 1361 EGRQVKGHFVTINN--QRYFLDGDGSETAPSRPV-TENNKWYVDGNGKLVKGAQVINGN 1417
```

RESULT 10

US-09-499-203-2

; Sequence 2, Application US/09499203

; Patent No. 6570065

; GENERAL INFORMATION:

; APPLICANT: KOSSMANN, Jens

; APPLICANT: WELSH, Thomas

; APPLICANT: OUANZ, Martin

; APPLICANT: KNUTH, Karola

; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternansucrase

; FILE REFERENCE: 147-196P

; CURRENT APPLICATION NUMBER: US/09/499,203

; CURRENT FILING DATE: 2000-02-08

; NUMBER OF SEQ ID NOS: 54

; SOFTWARE: Patent In Ver. 2.1

; SEQ ID NO 2

; LENGTH: 2057

; TYPE: PRT

; ORGANISM: Leuconostoc mesenteroides

US-09-499-203-2

Query Match 35.0%; Score 2501.5; DB 2; Length 2057;

Best Local Similarity 43.8%; Pred. No. 1.2e-173;

Matches 586; Conservative 167; Mismatches 397; Indels 189; Gaps 41;

QY 138 GKEVPANN-----NTAHPKTVTDKIVIPKPKIGKLA-----QPSSLSODDIAALGN 183

Db 281 GGFQDVNNKRIYFAPNTGNNAVANTEII-----NGKQGRDANGNQVKNAFSKDVA--GN 332

QY 184 -----VKNIRKVNKGYYYKEDGTLOKNYALNINGKTFPFD-ETGALSNNTLPSK 232

Db 333 TFYFDANGVMLTGLQTSIGKTYLDEQHLRKNYAGTFNNQPMYFDADTGAKTAIEYQF 392

QY 233 KGNITNNDNTNSFAQYNOVYSTDVANFEHVDHYLPAESWYRPKYILKDGKTWTQSTEKDF 292

Db 393 DQGLVSQSNENT--PHNAAKSYDKSSPENVDGYLTADTWYRPTDILKNGDTWTASTETDM 450

QY 293 RPLMTWMPDDEBTQBOYVYNNNAO-LGIHQTYNTATSPQLNLAQTOTKLEBKITABK 351

Db 451 RPLMTWMPDQOTQANYLNFMSSKGLGTTTTYTAATSQKTLNDAAFAVQTALIEQISLKK 510

QY 352 NTNMLROTISAFVKTSQSAWNSDSE-KPFD--DHLOKQGALLYSNNSKLTQSAQS-NVRIIL 407

Db 511 STEWLRLDAIDSFVKTOANWKNQTEDEAFDGLQWLOGGFAYQDDSHRTPNTDTSNNRKL 570

QY 408 RPTNTQTKGKDPRTADRTIGGYEFLLANDVNSNPVQABQLNWLHFMFNFGNIYANDP 467

Db 571 RQPINIDGSKD--TTDQK--GSEFLLANDIDNSNPIVQAEQLNWLHLMNFGSITGND 625

QY 468 DANFDSIRVDVNDVADLLOIAGDYLAKAAGIHKNDKAANDHLSILEAWSYNDTPYLD 527

Db 626 NANFDGIRVDVNDVADLLOIAGDYLAKAAGIHKNDKAANDHLSILEAWSYNDTPYLD 685


```
Db 780 NTPNLKLVNDKVVLLHMGAAHKNQYRAAVALTTTDDGVINTSDQAP--VAMTDENGDL 837
Qy 799 FTA-----AD--IKGYANPOVSGYLGWVVPVGAADQDVRVAASTAPSTDGKSVH 846
Db 838 LSSHNLVWNGKEADTAVQYANPDVSGYLAVVVPVGSADNQDARTAPSTEKNSGNSAYR 897
Qy 847 QNAALDSRVMEGFSNFOAPATKKEEYTNVVIKAVNDKFAEWGVTDFEMAPQVVSSTDG 906
Db 898 TNAAFDSNVIFEAFSNFVYPTFKESERANVRIAQNAADFFASLGFTSFEMAPQVNSKORT 957
Qy 907 FLDSVTQNGYAFDRYDLGLGSKPNKYGTADDLVKAIKALHSKGIKVMADWVDPQYALPE 966
Db 958 FLDSITDNGYAFDRYDLGSEBENKGTDEDLRNAIQALHKAQLQVMAWVDPQIYNLPG 1017
Qy 967 KEVVTATRVKDYGTVPVAGSQIKNTLYVVDGKSGKQQAQYGGAFLEELQAKYPELFARK 1026
Db 1018 KEVATVTRVDDRGVNWKDAIINNLYVNVN-TIGGGEYQKYGGAFLDKLQKLYPEIFTK 1076
Qy 1027 QISTGVPMPSVKIKOWSAKYFNGTNIILRGAGYVLKQATNTYFSLVSDNTFLPKSLV- 1085
Db 1077 QVSTGVAIDPSQKITEWSAKYFNGTNIILHRSGYVLKADGGQYVNLGTTTKOFLPQLTG 1136
Qy 1086 NPNHGTSSSVTLGLVFDGKGVYVYSTSGNOAKNAFI--SLGNWYVFDNNGYVMTGAQSN 1143
Db 1137 EKKQNEGVK--NDG-NYFYDLAGNVKVTIEDSVG-NWYFDDQDKGVENKHFVD 1192
Qy 1144 -----GANYFSLNGIOLRNAIYDNGNKVLSYGNDRRYENGYYLFGQWRYFQNG-GI 1196
Db 1193 VDSYGEKGYFFLKGVSFRGLVQT-----DNGTY-----YFDNYCK 1230
Qy 1197 MAVGLTRVHGAVQY-FDASG-----FOAKGQFITTAD-GKL 1230
Db 1231 MYRNOTINAGAMIYTLDENGKLIKASYNDAEYPTSTDVGKM 1272

RESULT 12
US-09-995-749A-2
; Sequence 2, Application US/09995749A
; Patent No. 6867026
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: B043388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1781
; TYPE: PRN
; ORGANISM: Lactobacillus reuteri
US-09-995-749A-2

Query Match 32.9%; Score 2347.5; DB 2; Length 1781;
Best Local Similarity 41.3%; Pred. No. 1.9e-162;
Matches 538; Conservative 181; Mismatches 408; Indels 175; Gaps 35;

Qy 29 TSLSSSLVKAUSTDRQOAVTESQASLVTTSFAAKETLTATDTSTATSQPTATVTDN 88
Db 549 SAVSGFDTTIKLTNDQYQALNGQLVLLRFSKAADG--NPSGDNTVTDQFSKNYATTG 606
Qy 89 VS-----TTNOS-----TWNTANTA-----NFVV 107
Db 607 FDYKVGNGVQEFSGWHATNSNDKDSQWIIVLVNGKEVKQLVNDTKEGAGFNNDVY 666
Qy 108 KPTTSEQATDNSDKIITTSKAVNR-----LTATGKFVFPANNNTAHPKTVT 154
```

```
Db 667 KVNPAIENSSMGFOGIITLPVTVKXNVQLVHRFSNDVKTGEGNVY-----DFW 716
Qy 155 DKIVPIKPKI---GKLKOPSSLSQDDIAALGNVKNIRKUNGKYYY-KEDGTLOKNYAL 209
Db 717 SELMPFVDSFQKNGPLKQ-----FGLQTINGOQYIDPTTGPQRKNFL 761
Qy 210 NINGKTFEEDTCALSNNTLPSK--KGNITNDNTINSFAQYNQVYSTDVANFHVHLYT 267
Db 762 QSGNWIYFDSDTGVGTNALELQFAKGTVSSNE---QYRNGNAAYSYDDKSIENVNGYLT 818
Qy 268 ASWTRPKVILKDGKTWTOSTEKDPRPLMTWHPDOETORQYVYNNAOLGIHQ-----Y 323
Db 819 ADWTRPKOILKDGTTWTDSTKETDMRPILMVWMPNTLTQAYLYNMKQHGNLPSALPFF 878
Qy 324 NTATSPQLNLAAOTIQTKEIKTAETAKTNMLRQTIISAFVKTQTSANWSDSEKPPDDHLQ 383
Db 879 NADADPAELNHSEIVQOQNIKRISETGNTDMLRMLHDFVTNNPMWNKNSNVFSGIQ 938
Qy 384 -KGALLYSNNSKLTSQANSYRILARTPTNQTCKKDPRTYADRTIGGYEFLANDVDSN 442
Db 939 FQGGFLKYENSDLTPVANSYRLLGRMPIN-----IKOQYRSGQEFLLANDIDSN 989
Qy 443 PVVQAEQLNHLPLMNFNIYANDPANFDSIRVDADVNDADLLQIAGDYLKAAGIHK 502
Db 990 PVVQAEQLNWLYYLLNFGTITANNDOANFDSVRVDAPDNIDADLMNIAQDYFNAAYGM-D 1048
Qy 503 NDKAANDHLSILLEANSYNDTPYLHDDGDNWINDNRLSLLSYLAKPLNQRSGMPLIT 562
Db 1049 SDAVSNKHINILEDWNHADPEYFNKIGNPQLTMDDTIKNSLNHLSDATN-RWGLDAIVH 1107
Qy 563 NSLVNRTDDNAETAAPVPSYFIRAHDSSEVQDLIRNIIRTEINPNVVG---YSFTTEIKK 619
Db 1108 QSLADRENNSTENVIPNYSFVRAHDNNSODOQNAIR-----DVTGKYHYHFFDEBQK 1162
Qy 620 AFEIYNKOLLATEKYYTHYNTALSYALLTNKSSVPRVYVYGMFTDDGGQYMAHKTINYEA 679
Db 1163 GIDAYIQONSTVKKYNYLNPASVAILLTNKTDTIPRVYVYDLYTDGGQYMEHQTRYDT 1222
Qy 680 IETLLKARKYVSGGOAMRNQOV--CNSEIITSVRYGKALKATDGTDRTRTSQVAVIEG 738
Db 1223 LTNLLKSRVYVAGGQSMQMTSVGGNNILTSVRYGKGAMTATDGTDETRTQIGVYVS 1282
Qy 739 NNPSLRLKASDRVVVNMGAHKNQYRPLLLTTDNGIKAYHSDQEAAGLVRTVNDRGELI 798
Db 1283 NTPNLKLVNDKVVLLHMGAAHKNQYRAAVALTTTDDGVINTSDQAP--VAMTDENGDL 1340
Qy 799 FTA-----AD--IKGYANPOVSGYLGWVVPVGAADQDVRVAASTAPSTDGKSVH 846
Db 1341 LSSHNLVWNGKEADTAVQYANPDVSGYLAVVVPVGSADNQDARTAPSTEKNSGNSAYR 1400
Qy 847 QNAALDSRVMEGFSNFOAPATKKEEYTNVVIKAVNDKFAEWGVTDFEMAPQVVSSTDG 906
Db 1401 TNAAFDSNVIFEAFSNFVYPTFKESERANVRIAQNAADFFASLGFTSFEMAPQVNSKORT 1460
Qy 907 FLDSVTQNGYAFDRYDLGLGSKPNKYGTADDLVKAIKALHSKGIKVMADWVDPQYALPE 966
Db 1461 FLDSITDNGYAFDRYDLGSEBENKGTDEDLRNAIQALHKAQLQVMAWVDPQIYNLPG 1520
Qy 967 KEVVTATRVKDYGTVPVAGSQIKNTLYVVDGKSGKQQAQYGGAFLEELQAKYPELFARK 1026
Db 1521 KEVATVTRVDDRGVNWKDAIINNLYVNVN-TIGGGEYQKYGGAFLDKLQKLYPEIFTK 1579
Qy 1027 QISTGVPMPSVKIKOWSAKYFNGTNIILRGAGYVLKQATNTYFSLVSDNTFLPKSLV- 1085
Db 1580 QVSTGVAIDPSQKITEWSAKYFNGTNIILHRSGYVLKADGGQYVNLGTTTKOFLPQLTG 1639
Qy 1086 NPNHGTSSSVTLGLVFDGKGVYVYSTSGNOAKNAFI--SLGNWYVFDNNGYVMTGAQSN 1143
Db 1640 EKKQNEGVK--NDG-NYFYDLAGNVKVTIEDSVG-NWYFDDQDKGVENKHFVD 1695
Qy 1144 -----GANYFSLNGIOLRNAIYDNGNKVLSYGNDRRYENGYYLFGQWRYFQNG-GI 1196
```


TYPE: PRT
ORGANISM: Streptococcus mutans
US-09-995-749A-10

Query Match 23.7% Score 1690.5; DB 2; Length 545;
Best Local Similarity 60.1%; Pred. No. 3.7e-115;
Matches 328; Conservative 84; Mismatches 115; Indels 19; Gaps 4;

QY 433 LLANDVDSNPVQABQLNHLFLMNFNGIYANDPDANFDSIRVDADVNDADLLQIAGD 492
Db 1 LLANDIDNSNPVQAEQLNHLVLMNYGSIIVANDPEANFDGVRVDADVNDADLLQIAGD 60

QY 493 YLKAAGIHKNDKANDHLSILEAWSYNDTPYLHDDGDMNMINNDRLRLSLYSLAKPLN 552
Db 61 YLKAHGVDKSEKNAINHLSILEAWSNDPQYNKDTKGALPDKLRLSLYALTRPLE 120

QY 553 Q-----RSGMNPILTNLVRNRTDNDNAETAAPSPYSFIRAHDSVQDLIRNIRTEIN 604
Db 121 KQASKNEIRSGLEPVITNSLNRSAEGKNSERMANYIFIRAHDSVQTVIAKIKAQIN 180

QY 605 PNWVGYSFTTEIKKAFIYNKDLATEKKYTHYNTALSYALLTNKSSVPRVYGDMEF 664
Db 181 PKTDGLTFLDELKQAFIYNEDMRQAKKKYQSNIPYALMLSNKDSITRLYYGDMYS 240

QY 665 DQGQYMAHKTINYAEITLLKARIKYVSGQAMRNOQVGNSE-----IITSVRYCK 715
Db 241 DQGQYMATKSPYYDAIDTLLKARIKYAAGQDMKITVEGDKSHMDWDYTVLTSVRYGT 300

QY 716 GALKATDTCGRTRISGVAVIEGNNPSLRLKASDRVVMGAAHKNQAYRPLLLTTDNGI 775
Db 301 GANEATDQGEATKQGMVITSSNPSLKLNQNDKVIYVNMGAHKNQYRPLLLTTDGL 360

QY 776 KAYHSDQEAAGLVRYTNRGELIFTAADIKG-YANPQVSGYLGVVVPVGAADQDVRVAA 834
Db 361 TSYTSDAAKSLYRKTNKDELVDASDIQGLYLPQVSG-LAVWVPVGAADQDVRVAA 419

QY 835 STAPSTDGKSVHONAAALDSRVNFEFSNFQAFATKKEYTNNVVIKQNDKFAEWGVTDFE 894
Db 420 SNKANATGVYESSALDSQLIYEGFSNFQDFVTKDSYTNKKIAQNVOLFCKSWGVTSE 479

QY 895 MAPQYVSSDGSFLDSVIONGYAFTDRYDLGISKNKYGTADDLVKAIKALHSGIKVMA 954
Db 480 MAPQYVSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSQQDMINAVKALHSGIQVIA 539

QY 955 DWVPDQ 960
Db 540 DWVPDQ 545

Search completed: February 11, 2006, 19:45:39
Job time : 35.1047 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 20:39:23 ; Search time 123.071 Seconds
(without alignments)
4668.162 Million cell updates/sec

Title: US-10-797-821-35
Perfect score: 7143
Sequence: 1 MEKKVRFKLKRVKRWTVTS.....VKVSVNTMILIPMKRFVIM 1375

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA Main:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pcp:*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pcp:*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pcp:*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pcp:*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pcp:*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pcp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7143	100.0	1375	3	US-09-740-274-4
2	7143	100.0	1375	4	US-10-383-930-35
3	7143	100.0	1375	5	US-10-797-821-35
4	5285	74.0	1475	3	US-09-740-274-2
5	5285	74.0	1475	4	US-10-383-930-34
6	5285	74.0	1475	5	US-10-797-821-34
7	4214.5	59.0	1590	4	US-10-383-930-37
8	4214.5	59.0	1590	5	US-10-797-821-37
9	3591.5	50.3	1430	3	US-09-740-274-6
10	3591.5	50.3	1430	4	US-10-383-930-36
11	3591.5	50.3	1430	5	US-10-797-821-36
12	3161.5	44.3	1554	4	US-10-383-930-38
13	3161.5	44.3	1554	5	US-10-797-821-38
14	3113	43.6	1497	5	US-10-484-218-18
15	3070	43.0	1518	4	US-10-383-930-40
16	3070	43.0	1518	5	US-10-797-821-40
17	2893.5	40.5	1365	4	US-10-383-930-39
18	2893.5	40.5	1365	5	US-10-797-821-39
19	2861	40.1	1595	5	US-10-484-218-20
20	2501.5	35.0	2057	4	US-10-417-280A-2
21	2410	33.7	1006	5	US-10-484-218-22
22	2383.5	33.4	1771	5	US-10-484-218-14
23	2383.5	33.4	1771	3	US-10-484-218-12
24	2347.5	32.9	1781	3	US-09-995-749A-2
25	1690.5	23.7	545	3	US-09-995-749A-10
26	1616	22.6	522	3	US-09-995-749A-11
27	1445.5	20.2	535	3	US-09-995-749A-13

28	1421	19.9	787	5	US-10-484-218-16	Sequence 16, Appl
29	1337	18.7	584	3	US-09-995-749A-12	Sequence 12, Appl
30	1273.5	17.8	525	5	US-10-484-218-23	Sequence 23, Appl
31	677.5	9.5	224	5	US-10-484-218-4	Sequence 4, Appl
32	673	9.4	223	5	US-10-484-218-6	Sequence 6, Appl
33	641	9.0	223	5	US-10-484-218-10	Sequence 10, Appl
34	585	8.2	221	5	US-10-484-218-2	Sequence 2, Appl
35	553	7.7	221	5	US-10-484-218-8	Sequence 8, Appl
36	312.5	4.4	2710	4	US-10-011-366-6	Sequence 6, Appl
37	312.5	4.4	2710	4	US-10-354-774-6	Sequence 6, Appl
38	312.5	4.4	2710	4	US-10-271-012-6	Sequence 6, Appl
39	312.5	4.4	2710	4	US-10-729-122-6	Sequence 6, Appl
40	312.5	4.4	2710	4	US-10-729-039-6	Sequence 6, Appl
41	312.5	4.4	2710	5	US-10-729-527-6	Sequence 6, Appl
42	312.5	4.4	2710	5	US-10-727-898-6	Sequence 6, Appl
43	312.5	4.4	2710	5	US-10-728-696-6	Sequence 6, Appl
44	312.5	4.4	2710	6	US-11-001-241-6	Sequence 6, Appl
45	307	4.3	726	4	US-10-282-122A-72033	Sequence 72033, A

ALIGNMENTS

RESULT 1
US-09-740-274-4
; Sequence 4, Application US/09740274
; Patent No. US20020031826A1
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-4

Query Match 100.0%; Score 7143; DB 3; Length 1375;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MEKKVRFKLKRVKRWTVTSIASAVVTLTSLSGSLVKADSTDDRQAVTESQASLVTTSE	60
Db	1	MEKKVRFKLKRVKRWTVTSIASAVVTLTSLSGSLVKADSTDDRQAVTESQASLVTTSE	60
QY	61	AAKETLTATDTSTATSQPTATVTDNVSTNQTNTANTANFVVKPTTSEAKTDN	120
Db	61	AAKETLTATDTSTATSQPTATVTDNVSTNQTNTANTANFVVKPTTSEAKTDN	120
QY	121	SDKIITTSKANRLTATGCKFVPANNNTAHPKTVTDKIPIKPKIGKLPKPSLSODDIAA	180
Db	121	SDKIITTSKANRLTATGCKFVPANNNTAHPKTVTDKIPIKPKIGKLPKPSLSODDIAA	180
QY	181	LGNVKNTRKVGKYYKEDGTLOKNVALNINGKTFFFDETCALSNNTLPSKKGNITNND	240

Db 181 LGNVKNIRKVGKYYKEDGTLQKNYALNINNGKTFPFDETCALSNNTLPSKKGNITNND 240
Qy 241 NTNSFAQYNQVSTVDVANFEHVDHYLTAEWSYRPKYILKDGKTWTQSTEKDPRPLMTW 300
Db 241 NTNSFAQYNQVSTVDVANFEHVDHYLTAEWSYRPKYILKDGKTWTQSTEKDPRPLMTW 300
Qy 301 PQDETQRYVYVMAQLGIHQTYNTATSPQLNLAAQTIOTKIEEKI TAEKNTNMLRQTI 360
Db 301 PQDETQRYVYVMAQLGIHQTYNTATSPQLNLAAQTIOTKIEEKI TAEKNTNMLRQTI 360
Qy 361 SAFVKTQSAWNSDSEKPPDDHLQKALLYSNNSKLTQSAWSYRILNRTPTNQTKGKDP 420
Db 361 SAFVKTQSAWNSDSEKPPDDHLQKALLYSNNSKLTQSAWSYRILNRTPTNQTKGKDP 420
Qy 421 YTADRTIGGYEFLANDVNSNPVQAEQLNLHFLMNFNTYANDPDANFDSIRDAVD 480
Db 421 YTADRTIGGYEFLANDVNSNPVQAEQLNLHFLMNFNTYANDPDANFDSIRDAVD 480
Qy 481 NVADALLQIAGDYKAAKGHKNDKAANDHLSILEAWSYNDTPYLHDDGDNMINDNRLR 540
Db 481 NVADALLQIAGDYKAAKGHKNDKAANDHLSILEAWSYNDTPYLHDDGDNMINDNRLR 540
Qy 541 LSLLYSLAKPLNQRSGMNPITNSLVNRTDDNAETAAPVSYFIRAHDSVODLIRNIIR 600
Db 541 LSLLYSLAKPLNQRSGMNPITNSLVNRTDDNAETAAPVSYFIRAHDSVODLIRNIIR 600
Qy 601 TEINPNVGYSTTBEIKAEFEIYNKDLATSEKTHYNTALSYALLTNKSSVPRVYG 660
Db 601 TEINPNVGYSTTBEIKAEFEIYNKDLATSEKTHYNTALSYALLTNKSSVPRVYG 660
Qy 661 DMFTDDGQYMAKHTINYEAEITLLKARIKYVSGQAMRNQOVGNSEIITSVRYGKALKA 720
Db 661 DMFTDDGQYMAKHTINYEAEITLLKARIKYVSGQAMRNQOVGNSEIITSVRYGKALKA 720
Qy 721 TDTGDRTRTSGVAVIEGNPNLSRLKASDRVVVNMGAHKNOAYRPLLLTTDNGIKAYHS 780
Db 721 TDTGDRTRTSGVAVIEGNPNLSRLKASDRVVVNMGAHKNOAYRPLLLTTDNGIKAYHS 780
Qy 781 DQEAAGLVRYTNDRGELIPTAADIKGYANPQVSGYLGWVPVGAADQDVRVAASTAPST 840
Db 781 DQEAAGLVRYTNDRGELIPTAADIKGYANPQVSGYLGWVPVGAADQDVRVAASTAPST 840
Qy 841 DGKSVHQNAALDSRVWFEGSFQAFATKKEEYTNVVIKAKNDKPAEKGVTDFENAPQV 900
Db 841 DGKSVHQNAALDSRVWFEGSFQAFATKKEEYTNVVIKAKNDKPAEKGVTDFENAPQV 900
Qy 901 SSTDGSFLDSVLQNGYAFTRDYDLGTSKPNKYGTADDLVKAIKALHSGIKVMADWVPDQ 960
Db 901 SSTDGSFLDSVLQNGYAFTRDYDLGTSKPNKYGTADDLVKAIKALHSGIKVMADWVPDQ 960
Qy 961 MYALPKEVVTATRVDKYGTVPVAGSQIKNTLVVDGKSSGKQQAQKYGGAFLLEELQAKYP 1020
Db 961 MYALPKEVVTATRVDKYGTVPVAGSQIKNTLVVDGKSSGKQQAQKYGGAFLLEELQAKYP 1020
Qy 1021 ELFPARKQISTGVPMDPSVKIKOWSAKYFNGTNILGRGAGYVLKDAQNTYFSLVSDNTP 1080
Db 1021 ELFPARKQISTGVPMDPSVKIKOWSAKYFNGTNILGRGAGYVLKDAQNTYFSLVSDNTP 1080
Qy 1081 PKSLVNPNHGTSSTVGLVFDGKGYVYYSYSGNQAKNAFISLGNWYFDDNNGYMVTGAQ 1140
Db 1081 PKSLVNPNHGTSSTVGLVFDGKGYVYYSYSGNQAKNAFISLGNWYFDDNNGYMVTGAQ 1140
Qy 1141 SINGANYTFLSNGIQLRNAIYONGNKVLSSYNGDGRYENGYYLFGQQWRYFQNGIMAVG 1200
Db 1141 SINGANYTFLSNGIQLRNAIYONGNKVLSSYNGDGRYENGYYLFGQQWRYFQNGIMAVG 1200
Qy 1201 LTRVHGAQYFDPASGFQAKGQFIITADGKLRVFDKSGNQISNR FVRNSKGEWFLFDHNG 1260
Db 1201 LTRVHGAQYFDPASGFQAKGQFIITADGKLRVFDKSGNQISNR FVRNSKGEWFLFDHNG 1260
Qy 1261 VAVTGTVTTFNGQRLYFKPKNVQAKGEFIRDANGYLRYDNPNSGNEVRNRFVNSKGEWFL 1320
Db 1261 VAVTGTVTTFNGQRLYFKPKNVQAKGEFIRDANGYLRYDNPNSGNEVRNRFVNSKGEWFL 1320

RESULT 2

US-10-383-930-35
; Sequence 35, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patent In version 3.2
; SEQ ID NO 35
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-35

Query Match 100.0%; Score 7143; DB 4; Length 1375;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MEKKVRFKLRKVKRWVTVSIASAVVTLTSLSGSLVKADSTDDRQOAVTESQASLVTTSE 60
Db 1 MEKKVRFKLRKVKRWVTVSIASAVVTLTSLSGSLVKADSTDDRQOAVTESQASLVTTSE 60
Qy 61 AAKETLTATDTSTATSATSOPTATVTDNVSTNQSTNTTANTANFVVKPTTTTSEQAKTDN 120
Db 61 AAKETLTATDTSTATSATSOPTATVTDNVSTNQSTNTTANTANFVVKPTTTTSEQAKTDN 120
Qy 121 SDKIITTSKAVNRLTATGKFPANNNTAHPKTVTDKIVIPKPKIGKLQPSLSQDDIAA 180
Db 121 SDKIITTSKAVNRLTATGKFPANNNTAHPKTVTDKIVIPKPKIGKLQPSLSQDDIAA 180
Qy 181 LGNVKNIRKVGKYYKEDGTLQKNYALNINNGKTFPFDETCALSNNTLPSKKGNITNND 240
Db 181 LGNVKNIRKVGKYYKEDGTLQKNYALNINNGKTFPFDETCALSNNTLPSKKGNITNND 240
Qy 241 NTNSFAQYNQVSTVDVANFEHVDHYLTAEWSYRPKYILKDGKTWTQSTEKDPRPLMTW 300
Db 241 NTNSFAQYNQVSTVDVANFEHVDHYLTAEWSYRPKYILKDGKTWTQSTEKDPRPLMTW 300
Qy 301 PQDETQRYVYVMAQLGIHQTYNTATSPQLNLAAQTIOTKIEEKI TAEKNTNMLRQTI 360
Db 301 PQDETQRYVYVMAQLGIHQTYNTATSPQLNLAAQTIOTKIEEKI TAEKNTNMLRQTI 360
Qy 361 SAFVKTQSAWNSDSEKPPDDHLQKALLYSNNSKLTQSAWSYRILNRTPTNQTKGKDP 420
Db 361 SAFVKTQSAWNSDSEKPPDDHLQKALLYSNNSKLTQSAWSYRILNRTPTNQTKGKDP 420
Qy 421 YTADRTIGGYEFLANDVNSNPVQAEQLNLHFLMNFNTYANDPDANFDSIRDAVD 480
Db 421 YTADRTIGGYEFLANDVNSNPVQAEQLNLHFLMNFNTYANDPDANFDSIRDAVD 480
Qy 481 NVADALLQIAGDYKAAKGHKNDKAANDHLSILEAWSYNDTPYLHDDGDNMINDNRLR 540
Db 481 NVADALLQIAGDYKAAKGHKNDKAANDHLSILEAWSYNDTPYLHDDGDNMINDNRLR 540
Qy 541 LSLLYSLAKPLNQRSGMNPITNSLVNRTDDNAETAAPVSYFIRAHDSVODLIRNIIR 600
Db 541 LSLLYSLAKPLNQRSGMNPITNSLVNRTDDNAETAAPVSYFIRAHDSVODLIRNIIR 600

601 TEINPNVVGYSFTTBEIKKAFIYNKOLLATEKKYTHYNTALSYALLLTKNSVPRVYIG 660
Db TEINPNVVGYSFTTBEIKKAFIYNKOLLATEKKYTHYNTALSYALLLTKNSVPRVYIG 660
661 DMFTDDGQYMAHKTINYEAIETLLKARIKYVSGQAMRNOQVGNSEIITSVRYGKALK 720
Db DMFTDDGQYMAHKTINYEAIETLLKARIKYVSGQAMRNOQVGNSEIITSVRYGKALK 720
721 TDGDRTRTTSQVAVIEGNNPSLRLKASDRVVVNMGAHKNAQYRPLLLTTDNGIKAYHS 780
Db TDGDRTRTTSQVAVIEGNNPSLRLKASDRVVVNMGAHKNAQYRPLLLTTDNGIKAYHS 780
781 DQEAAGLVRYTNDRGELIFTAADIKGYANPOVSGYLGWVPVGAADQDVRVAASTPST 840
Db DQEAAGLVRYTNDRGELIFTAADIKGYANPOVSGYLGWVPVGAADQDVRVAASTPST 840
841 DGKSVHQNAAALDSRVNMFEGSFNFOAFATKKEEYTNVVIKNVDKPAEWGVTDFEMAPQYV 900
Db DGKSVHQNAAALDSRVNMFEGSFNFOAFATKKEEYTNVVIKNVDKPAEWGVTDFEMAPQYV 900
901 SSTDGSFLDSVIQNGYAFTRDYLGLISKPNKYGTADDLVKAIKALHSKGIKVMADWVPDQ 960
Db SSTDGSFLDSVIQNGYAFTRDYLGLISKPNKYGTADDLVKAIKALHSKGIKVMADWVPDQ 960
961 MYALPEKEVVTATRVVDKYGTPVAGSQIKNTLYYVDGKSSGKQQAQYGGAFLEELQAKYP 1020
Db MYALPEKEVVTATRVVDKYGTPVAGSQIKNTLYYVDGKSSGKQQAQYGGAFLEELQAKYP 1020
1021 ELFPARKQISTGVPMDSVKIKOWSAKYFNGTNILGRGAGVYLKQDQATNTYFSLVSDNTFL 1080
Db ELFPARKQISTGVPMDSVKIKOWSAKYFNGTNILGRGAGVYLKQDQATNTYFSLVSDNTFL 1080
1081 PKSLVNPNHGTSSTGLVFDGKGYVYYSYSGNQAKNAFISLGNWYYPDNNGYMTGAQ 1140
Db PKSLVNPNHGTSSTGLVFDGKGYVYYSYSGNQAKNAFISLGNWYYPDNNGYMTGAQ 1140
1141 SINGANYPLSNGIQLRNAIYDNGNKVLSYNGDGRYENGYYLFCQWRIFYQNGIMAVG 1200
Db SINGANYPLSNGIQLRNAIYDNGNKVLSYNGDGRYENGYYLFCQWRIFYQNGIMAVG 1200
1201 LTRVHGAVOYFASGFOAKGQFTTADGKLRVFDKRDGSGNOISNRFRVNSKGEMFLPDHNG 1260
Db LTRVHGAVOYFASGFOAKGQFTTADGKLRVFDKRDGSGNOISNRFRVNSKGEMFLPDHNG 1260
1261 VAVTGTVTFNGORLPKPNVGQAKGEFIRDANGYLRYYDPNPSGNEVRNRFVNSKGEMFL 1320
Db VAVTGTVTFNGORLPKPNVGQAKGEFIRDANGYLRYYDPNPSGNEVRNRFVNSKGEMFL 1320
1321 FDHNGIATVGARVNGHASILSMVFRLESSLQSVKVVSNMTILIPMKFVIVM 1375
Db FDHNGIATVGARVNGHASILSMVFRLESSLQSVKVVSNMTILIPMKFVIVM 1375

RESULT 3
US-10-797-821-35

; Sequence 35, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550

; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 35
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-35

Query Match 100.0%; Score 7143; DB 5; Length 1375;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEKKVRPKRKVKRWVTVSIASAVVTLTSLGSLVKADSTDDRQQAQVTSQASLVTTSE 60
Db 1 MEKKVRPKRKVKRWVTVSIASAVVTLTSLGSLVKADSTDDRQQAQVTSQASLVTTSE 60
61 AAKETLTATDTSTATSATSQPTATVTDVNTNQSTNTTANTANFVVKPTTSEQAKTDN 120
Db 61 AAKETLTATDTSTATSATSQPTATVTDVNTNQSTNTTANTANFVVKPTTSEQAKTDN 120
121 SKLIITTSKAVNRLTATGKFPVANNNTAHPKTVTDKI VPKPKIGLKPSSLSQDDIAA 180
Db 121 SKLIITTSKAVNRLTATGKFPVANNNTAHPKTVTDKI VPKPKIGLKPSSLSQDDIAA 180
181 LGNVNIRKVGKYYKEDGTLQKNYALNINNGKTFPDETGALSNTLPSKKGNTNND 240
Db 181 LGNVNIRKVGKYYKEDGTLQKNYALNINNGKTFPDETGALSNTLPSKKGNTNND 240
241 NTNSFAQYNQVSTVDVANFEHVDHYLTAEHWPRPKYILKDGKTWTQSTKDFRPLLMTW 300
Db 241 NTNSFAQYNQVSTVDVANFEHVDHYLTAEHWPRPKYILKDGKTWTQSTKDFRPLLMTW 300
301 PQDETQRQVYVNMNAQLGIHOTYNTATSPQLNLAQTTIQTKEEKITAEKNTNWLRTI 360
Db 301 PQDETQRQVYVNMNAQLGIHOTYNTATSPQLNLAQTTIQTKEEKITAEKNTNWLRTI 360
361 SAFVKTOSAWNSDSKPKDDHLQKALLYSNNSKLTQSANSNRYILNRPPTQTKGKDP 420
Db 361 SAFVKTOSAWNSDSKPKDDHLQKALLYSNNSKLTQSANSNRYILNRPPTQTKGKDP 420
421 YTADRTIGGYEFLANDVNSNPVQAEQNLNHLFNMFGNIYANDPDANFDSIRVDADV 480
Db 421 YTADRTIGGYEFLANDVNSNPVQAEQNLNHLFNMFGNIYANDPDANFDSIRVDADV 480
481 NYDADLLQIAGDYLKAAKGIHKNDKAAANDHLSILEAWSYNDTPYLHDDGDNINMDNRL 540
QY 481 NYDADLLQIAGDYLKAAKGIHKNDKAAANDHLSILEAWSYNDTPYLHDDGDNINMDNRL 540
541 LSLLYSLAKPLNORSQMNPLITNSLVNRTDDNAETAAPVPSYFIRAHDSQVODLIRNIIR 600
Db 541 LSLLYSLAKPLNORSQMNPLITNSLVNRTDDNAETAAPVPSYFIRAHDSQVODLIRNIIR 600
601 TEINPNVVGYSFTTBEIKKAFIYNKOLLATEKKYTHYNTALSYALLLTKNSVPRVYIG 660
Db 601 TEINPNVVGYSFTTBEIKKAFIYNKOLLATEKKYTHYNTALSYALLLTKNSVPRVYIG 660
661 DMFTDDGQYMAHKTINYEAIETLLKARIKYVSGQAMRNOQVGNSEIITSVRYGKALK 720
QY 661 DMFTDDGQYMAHKTINYEAIETLLKARIKYVSGQAMRNOQVGNSEIITSVRYGKALK 720
721 TDGDRTRTTSQVAVIEGNNPSLRLKASDRVVVNMGAHKNAQYRPLLLTTDNGIKAYHS 780
Db 721 TDGDRTRTTSQVAVIEGNNPSLRLKASDRVVVNMGAHKNAQYRPLLLTTDNGIKAYHS 780
781 DQEAAGLVRYTNDRGELIFTAADIKGYANPOVSGYLGWVPVGAADQDVRVAASTPST 840
Db 781 DQEAAGLVRYTNDRGELIFTAADIKGYANPOVSGYLGWVPVGAADQDVRVAASTPST 840
841 DGKSVHQNAAALDSRVNMFEGSFNFOAFATKKEEYTNVVIKNVDKPAEWGVTDFEMAPQYV 900


```
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797, 821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383, 930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363, 209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402, 483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290, 049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081, 550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115, 142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 34
; LENGTH: 1475
; TYPE: PR1
; ORGANISM: Streptococcus mutans
US-10-797-821-34

Query Match          74.0%; Score 5285; DB 5; Length 1475;
Best Local Similarity 75.9%; Pred. No. 1.1e-296;
Matches 1031; Conservative 101; Mismatches 169; Indels 58; Gaps 14;

QY 1 MEKKVRFKLVKRWVTSIASAVVLTLSGSLVKADSTDDROQAVTESQASLVTTSE 60
DB 1 MDKVRVYKLVKRWVTSVASAVMTLTLSGGLVKADSNESQISNDSNTSVVTANE 60
QY 61 AKKELTATDSTATSAFSQPTA--TVTDNSTTQSTNTTANTANFVVKPTTSEQAKT 118
DB 61 -----ESNVITEATSKQEAASQTNHTVTTSSSTSVVNPKEVSNPYTVGETA-- 109
QY 119 DNSDKIITTSKAVNLATGKVPANNTAHPKTVTDKIVPKIKGLKQPSLSQDDI 178
DB 110 SNGEKLQNTTVDKTSSEA-----AANNISKOTTEAD-----TDVIDSN 149
QY 179 AA----LGNVRNIRKNGVYKYKEDGTLQKNYALNINNGKTFPFDGTGALSNNTLPSKGG 234
DB 150 AANLQILEKLPNVKEIDGKYVYDNGKVRNFTLIADKILHFDGTGAYTDTSIDTVNK 209
QY 235 NITNDNTNSPAQYNOVSTVDVANEHVDHYLTAEISWRPKYILKDGKTWTQSTSEKDRP 294
DB 210 DIVTT-RSNLYKYNQVYDRSAQSFEHVDHYLTAEISWRPKYILKDGKTWTQSTSEKDRP 268
QY 295 LLMTWPPQETQROVYVYVYNAQLGTHQVNTATSPQLNLAQTITQKIEEKITAEKNTN 354
DB 269 LLMTWPPQETQROVYVYVYVYNAQLGINKTYDDTSNQLNLNAAATIQAKIEAKITLTKNTD 328
QY 355 WLQRTISAFVKTQSAWNSDSEKFPDDHLQKGLLYSNNSKLTQSAQNSYRILNRTPTNQ 414
DB 329 WLQRTISAFVKTQSAWNSDSEKFPDDHLQNGAVLYDNEGKLTPTYANSYRILNRTPTNQ 388
QY 415 GKQDPYTRADRTIGYEFLANDVNSNPVQAEQLNLHLFLMFGNIYANDPDANFDSI 474
DB 389 GKQDPYTRADNTIGYEFLANDVNSNPVQAEQLNLHLFLMFGNIYANDPDANFDSI 448
QY 475 RVDADVNDVADLLQIAGDYLKAAGIKHNDKAANDHLSILEAWSYNDPTPYLHDDGDNMIN 534
DB 449 RVDADVNDVADLLQIAGDYLKAAGIKHNDKAANDHLSILEAWSYNDPTPYLHDDGDNMIN 508
QY 535 MDNRLRLSLLSLAKPLNQRGMNPLITNSLVNRTDDNAETAAPVPSYFIRAHDSVQDL 594
DB 509 MDNKLRLSLLSLAKPLNQRGMNPLITNSLVNRTDDNAETAAPVPSYFIRAHDSVQDL 568
QY 595 IRNIIRTEINPNVGYSTTEIEIKAPFIYNKOLLATEKKYTHYNTALSYALLNKKSSV 654
DB 569 TADIIKAEINPNVGYSTTEIEIKAPFIYNKOLLATEKKYTHYNTALSYALLNKKSSV 628
QY 655 PRVYVYGDFTDDGOVMAHKTINYEAEITLLKARIKIVYSGGQAMRNQOVGNSBIITSVRYG 714
DB 655 PRVYVYGDFTDDGOVMAHKTINYEAEITLLKARIKIVYSGGQAMRNQOVGNSBIITSVRYG 688
DB 688 PRVYVYGDFTDDGOVMAHKTINYEAEITLLKARIKIVYSGGQAMRNQOVGNSBIITSVRYG 688
QY 715 KGALKATDTGDRTRTTSVAVIEGNNPSLRLLKASDRVVVNMGAHKNQAYRPLLLTTDNG 774
DB 689 KGALKATDTGDRTRTTSVAVIEGNNPSLRLLKASDRVVVNMGAHKNQAYRPLLLTTDNG 748
QY 775 IKAYHSDQEAAGLVRYTNDRGELIFTAADIKGYANPOVSGYLVGVVVPVGAALIKWFALE 832
DB 749 IKAYHSDQEAAGLVRYTNDRGELIFTAADIKGYANPOVSGYLVGVVVPVGAALIKWFALE 808
QY 833 AASTAPSTDGKSVHQNAAALDSRVMPFSGNFQAFATKKBEYTNVVIKKNVDKFAEWGVD 892
DB 809 A---RPHQOMASVHQNAAALDSRVMPFSGNFQAFATKKBEYTNVVIKKNVDKFAEWGVD 865
QY 893 FEMAPQYVSSSTDGSLDSVIONGYAFTDRYDLGISKPNKYGTADDLVKAIKALHSGKIGV 952
DB 866 FEMAPQYVSSSTDGSLDSVIONGYAFTDRYDLGISKPNKYGTADDLVKAIKALHSGKIGV 925
QY 953 MADWVPDQMYALPEKEVVTATRVKYGTPVAGSQIKNTLYVVDGKSSGKDDQAKYGGAPL 1012
DB 926 MADWVPDQMYALPEKEVVTATRVKYGTPVAGSQIKNTLYVVDGKSSGKDDQAKYGGAPL 985
QY 1013 EELQAKYPELFAKQISTGVPMDFSVKIKQWSAKYFNGTNIILGRGAGYVVLKQATNTYFS 1072
DB 986 EELQAKYPELFAKQISTGVPMDFSVKIKQWSAKYFNGTNIILGRGAGYVVLKQATNTYFS 1045
QY 1073 LVSDN---TPIPKSLVNPNGHTSSVTGLVFDGKGYVYVYSTSGNOAKNAFISLGNWYVF 1129
DB 1046 -ISDNKEINFLPKTLN-----QDSQGVGFSYDGKGYVYVYSTSGYQAKNTFISEGDKWYVF 1099
QY 1130 DNGVMVTGAOSINGANYFELNSGILQRLNAYLDNGKNVLSYVYDNGRVEYNGYILF-GQQ 1188
DB 1100 DNGVMVTGAOSINGANYFELNSGILQRLNAYLDNGKNVLSYVYDNGRVEYNGYILF-GQQ 1159
QY 1189 WRYFQNGIMAYGLTRVHGAVQYFDASGFAQKGFITTDGKLYRFDSDSGNQISNRFVN 1248
DB 1160 WRHFNNGEMSVGLTVIDGQVYFDEMVGQAKGKFTVTADGKIRYFDKQSGNMYRNFEN 1219
QY 1249 SKGEWFLFDHNGVATGVTFTNGQRLYFKPNQVQAKGEFIRDANGYLYRYPNSGNEVN 1308
DB 1220 BEGKWLVLGEDGAAVGTSGTNGHLYFRANGVQVKGFEVTDHGHGRISYVDGNSGDIIRN 1279
QY 1309 RPYRNSKGEWFLFDHNGVATGVTGARVNVGHASILSLMVF 1347
DB 1280 RPYRNSKGEWFLFDHNGVATGVTGARVNVGHASILSLMVF 1313

RESULT 7
US-10-383-930-37
; Sequence 37, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 37
; LENGTH: 1590
; TYPE: PR1
; ORGANISM: Streptococcus sobrinus
US-10-383-930-37

Query Match          59.0%; Score 4214.5; DB 4; Length 1590;
Best Local Similarity 60.3%; Pred. No. 1e-234;
Matches 811; Conservative 186; Mismatches 296; Indels 53; Gaps 14;
```


RESULT 13
US-10-797-821-38
; Sequence 38, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: 2004-03-09
; PRIOR FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 38
; LENGTH: 1554
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-797-821-38

Query Match 44.3%; Score 3161.5; DB 5; Length 1554;
Best Local Similarity 48.5%; Pred. No. 7.8e-174;
Matches 662; Conservative 223; Mismatches 400; Indels 81; Gaps 31;

QY 1 MEKKVREKLKVKRWVTVSIASAVVTLTSLGSLVKADSTDRQAVTESQASI----- 55
DB 1 MEKKVHTKLKVKRWVTVSIASAVVTLTSLGSLVKADSTDRQAVTESQASI----- 55

QY 56 ---VTTSEAAKTLTATDTSTATSQTATVTDNVSTNTQSTNTTANFVVKPTTT 112
DB 57 QDQATTNDANTWT-DTDTADQSANNTQDQAG-SQSNQDQAKQDNTANT-----DRNQ 108

QY 113 SEQAKTNSDKLIITTSKAVNRLTATGKFPVANNNTAHPKTVTDKIVPKIKGUKLQPS 172
DB 109 ADNSQTDN-----NQATDQATS-----PATDGTSVQRDAANVATAADQEG--QTAP 153

QY 173 LSQDDIAALGNVKNIRKYNKYKYYKEDGTLOKNVALNINGKTFPPD-ETGALSNTLPS 231
DB 154 SEQEKSAAL-SLDNVKLLDGKYYVQADGSKYKKNFAITVNGQMLYFSDTGALSTSTYS 212

QY 232 KKGNTTNDNTNSFAQYNQVSTVDVANFEHVDHYLTAEWSYRPKYILKDGKTWTQSTBK 291
DB 213 FSQGTN-LVDDFSSHKNAYDSTAKSFELVNGYLTANSWYRPAIGLRNGQTMESNEND 270

QY 292 FRPLMTWMPDQETQRYVYNNYNAOLGIHQTYNT-ATSPQLNLAAQTIQKIEKITA 350
DB 271 LRPLVLMWMPDKDTQVAVYNNYNNKYLANSATEVTNETSQVDLKEAQSIQTQIEKITS 330

QY 351 KNTNWLRTQISAFVKTQSAWNSDSKPPD-DHLQKGLLYSNNSKLTQSAWNSYRI 408
DB 331 NSTQWLRTAMEAFVAAQFPKNMSTEN-FNKGHQLQGGLLYT-NSDLTFPWNDSYRLN 388

QY 409 TPTNQTGKDPRTYADRITGGYEFLLANDVDSNPNVQAEQNLNLHFLNFGNIYANDPD 468
DB 389 TPTQDQGT-KYFTEGEGGYEFLLSNDVDSNPNVQAEQNLNLHFLNFGNIYANDPD 446

QY 469 ANFDGIRVDVNDVADLLQIAGDYLKAAKGHKNDKKAANDHLSILEAWSYNDTPLYHDD 528
DB 447 ANFDGIRVDVNDVADLLQVYSNYFKONYKVTSDSEANALAHISILEAWSLNDQYNE 506

QY 529 GDNMNMNRLRLSLYLAKPLNQSRSGMNPILITNSL-VNRTDDNAETAAPVPSYFIRAH 587

Db 507 NCTALSIDNSSRLTSLAVLTQPCQRIDLSNLSVSKERANDTAYGDTITFTYFVRAH 566
QY 588 DSEVODLIRNIIRTEINPNVGVSTFTTEEIKKAFIYNKDLLEKKYTHYNTALSAYALL 647
Db 567 DSEVQTVIAKIVKEKIDTNSDGYTFTLDQLKDAFKIYNEDMAKVNKTTHYNTIPAYALL 626
QY 648 LTNKSSVRVYVYVDMFTDDGQYMAHKTINYEAIEFLKARIKYVSGGQAMRNQOV-GNSE 706
Db 637 LSNMESVRVYVYVYDLYDDGQYMAKSPYDAIATMLQGRYAYVSGGSEEVHVKVGNNG 686
QY 707 IITSVRYGKALKATDT-GDRTTRTSGVAVTEGNNPSRLKASDRVVMGAAHKNQAYR 765
Db 687 ILSSVRYGQDLMSADDTQGTDLSTRSGLVTLVNSDPNLDL-GGDSLTVMNGRAHQA 745
QY 766 PLLLTNDNGIKAYHSDQBAAGLVRYTNDRGELIFTAADIKGYANPOVGYLVGVVPGAA 825
Db 746 PLLILGTDGVQSYLKDS-D-TNIVKYTDANGNLTFADDIKGYSTVDMSGYLAVVVPVGA 804
QY 826 ADQVRAASTAPSTDGKSVHQNALDSRVNFEGESNFQAFATKKEEYTNVVIKNVDF 885
Db 805 DQDQVRAADTNQKADGSKSLKTSAAALDSQVIEGFSNFQDFANNDADYTNKKIABNAD 864
QY 886 AEWGVTDFEMAPQVYSTDSGSLDSVIQNGYAFTRDYDLGISKPNKYGTADDLVKAIKAL 945
Db 865 KKLGITSEMAPQVYSATDSGSLDSIIQNGYAFSDRYDLMSKNNKYGSKDDLALAKAL 924
QY 946 HSKGIKVMADVVPDQMYALPEKEVVVATRVKYGTPPVAGSQIKNTLYVVDGKSGKQQA 1005
Db 925 HANGIQAIAADVVPDQIYQLPGEBVVVTAKTNSYGNPTPDAYINNALYATNTKSSGSDY 984
QY 1006 KYGGAFLBELQAKYPELPARKQISTGVPMDSVKIKOWSAKYFNGTNIILGRGAGVYLKQ 1065
Db 985 QYGGAFDELKAKYPMDFVTNMISTGKPIDPSTKIKOWEAKYFNTNVLGKAGVYLSDD 1044
QY 1066 ATNTYFSLVSDNTFLPKSLVNPNGHSTSSVGLVDFGKGYVYVYSGNOAKNAFISLGN 1125
Db 1045 ATGKFTVNGEDFLPASFT-----GDQNAKTGYFDGTGMAYYSTSGKAVNSFIYEGH 1100
QY 1126 WYFDNNQYMTGA-QSINGANYPLSNGIQLRNAIYDNGNKVLSYDNGDRRYENGYVL 1184
Db 1101 YYYFDKDHMTGYSYKABDGDNDYFLPNGIQMRDAIYQDAQNSYYYGRTGLYK----- 1155
QY 1185 FGQW-----RYFO-NGINAVGLTRVHGAVQVDFDASGFOAKQOFITADGK 1229
Db 1156 -GDWNPYFVDPNNANKTVRYFDPANNMAIGYRNMYGQTYYPDENGFQAKGOLLTDDKG- 1213
QY 1230 LRYFDRDSNQISNRFRVNSKGEWFLFDHNGVAVTGTVTTFNGQRLYFKP-NGVQAKGEFI 1288
Db 1214 THYFDEDNGAMAKNFV-NUGDDWYMDGNGNAVKGOYFVNNQILLYFPETGVQVKGQFI 1272
QY 1289 RDANGLYRYDPNCGNEVRNFRVNSKGEWFLFDHNGIAVTGARVV 1334
Db 1273 TDAQRTSYDANSALKSSGFFTPNGSDWY-YAENGYYVYKGFQV 1317

RESULT 14
US-10-484-218-18
; Sequence 18, Application US/10484218
; Publication No. US20050059633A1
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; TITLE OF INVENTION: LACTIC ACID BACTERIA
; FILE REFERENCE: 2001-1316
; CURRENT APPLICATION NUMBER: US/10/484,218
; CURRENT FILING DATE: 2004-01-20
; PRIOR APPLICATION NUMBER: PCT/NL02/00495
; PRIOR FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: EP 01202752.0
; PRIOR FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: EP 01202841.1
; PRIOR FILING DATE: 2001-07-25

1

Matches 662; Conservative 223; Mismatches 410; Indels 136; Gaps 34;

```
QY 1 MEKKVRFKLRKVKRWVTVSIIAS-AVWTL---TSJSGSLVKADSTDDRQQAUTESQ---- 52
Db 1 MENKHYKHLKVKQKQWTVIAVASALATVLGSLSVTSSVSADETQDK--TVTQNSGTT 58
QY 53 ASLAVTSSAAKETLTATDTSTATSATSQPTATVTDNVSTTN-QST---NTTANTANFVVK 108
Db 59 ASLVTSPPEATKADKRNTKEADVLTPAKETNAVETATTTNTQATAEAATATTATADVAVA 118
QY 109 PTTTSEQAKTNSDKIIITTSKAVNRLTATGKFPVANNNTAHPKTVDTKIVPKIKGLK 168
Db 119 AVPNKEAVVTTDA-PAVYTEKAEQ-----PATYKAEVVNT-----EVK 156
QY 169 QP-SSLSQDDIAALGNVKNIRKNGKYYYKEDGTLOKNYALNINGTFFFDDETGALSNN 227
Db 157 APEAALKDSEVEAALSILKNIDGKYYYVNEDESHKENFAITVNGQLLYFGKDGALTSS 216
QY 228 TLPSSKGNITNNDNTNSFAQYNOVYSTDVANEHVDHYLTABSWYRPKVIKDKGTWTQS 287
Db 217 STYSFTPGTNN--IVDGFSSINRRAYDSEASEPFLIDGYLTADSWYRPASIIKDGVTWQAS 274
QY 288 TEKDFRPLMTWPPQETQROQVYNTMNAOLGIHQTYNTATSPLQNLAAQTQTKIEBK 347
Db 275 TAEDFRPLLMWPNVDVQVNYLNTMSKVFNLDAKYSSTDQETLKVAAKDIQIKIEQKI 334
QY 348 TAEKVTNMLRQTSAPVKTQSAWNSDSEK-----PFDDHLQKALLYSNNSKUTSQAANSY 403
Db 335 QAEKSTQWLRETISAPVKTPQWNKETENYSKGGEDHQQGALLYVNDNR-TPWANSY 393
QY 404 RLINPTPTNQTG-----KKDPRTADRTIGGYEFLANDNDVNSNPVQAEQLNMLH 454
Db 394 RLNLNTATNQTGTIDKSLDEQSDPNH-----MGGFDFLLANDVLSNPVQAEQLNQIH 448
QY 455 FLNPFNGIYANDPDANFDSIRDAVDNDVDADLLQAGDYLKAAGKGIHKNDKAANDHLSIL 514
Db 449 YLNNWGSIVMGDKANFDFGIRDAVDNDVDADMLQLYTNYFREYYGVNKEANALAHISVL 508
QY 515 EAWSYNDTPYLHDDGDNMINMDNRLSLLSYLAKPLNQRS--GMNPLITNSLVNRTDINA 573
Db 509 EAWSLNDNHNDKTDGAALAMENKQRLALLFSLAKPIKERTPAVSPLYNNTF-NTTORDE 567
QY 574 ET-----AAPSYSFIRAHDSVQDLIRNIIRTEI 603
Db 568 KTDWINKGSKAYNBDGTVKQSTIGKYNEKYDASGNYVFIKADNNVQDIIAEIHKKEI 627
QY 604 NPNVVGYSFTTEIKKAFIYNKDLATEKKTHTYNTALSYALLTNKSSVPRVYVGDMF 663
Db 628 NPKSDGFTITDAEMKQAFIYNKDLMLSSDKKYLNNIPAAVAVMLQNMETITRVYVGDL 687
QY 664 TDDGQYMAHKTINYBAIETLLKARIKYSGGQAMRNQ-----QVGN-----EIT 709
Db 688 TDDGHYMETKSPYDITVNLMSRIKYSGGQARSYMLPTDGMKDNDSVELYRTNEVYT 747
QY 710 SVRYGKGALKATDT-GDRTRTSYGAVIIEGNNPSRLKASDRVVNVMGAHKNOAVRPL 768
Db 748 SVRYGKIDINTDTSKYSRTSGQVTLVANNPKLNLDQSAKLNVEMGIHANQYIRALI 807
QY 769 LPTDNDIKAYHSDQE--AAGLVRYTNDRGELIFTAADIKGYANPOVSGYLVVVPVGA 826
Db 808 VGTADGINKFTSDADAIAGYVYKETSNGVLTFGANDIKGYETFDMSGFVAVVPVGSAD 867
QY 827 DQDVRAASTAPSTGK-SVHQNAAALDSRMPEGFSNFQAF--ATKKEEYTNVVIKQND 883
Db 868 NQDIRVAPSTEAKEGELTLKATEAYDSOLIEYEGFSNFQTI PDGSDPSVYTNRKIAENVD 927
QY 884 KFAEKGVTDFENAPVYVSDGSLDSVIONGYAFTDRYDLGISKENKYGTADDLVKAIK 943
Db 928 LFKSGVGTFSFEMAPVFSADGDTFLDSVIQNGYAFADRYDLAWSKNNKYGSKEDLRDAK 987
QY 944 ALHSGIKVMADVPDQMFALPEKEVVTATRVKYGTPVAGSQIKNTLYVVDGKSSGKDQ 1003
Db 988 ALHKAGIQAIADVPDQIYQLPGKEVVTATRTDAGAKIADAIIDHSLYVANSKSSGKDY 1047
```

```
QY 1004 QAKYGGAFLERLOAKYPPELPARKQISTGVPMDPSVKIKQWSAKYFNGTNIILRGAGYVLK 1063
Db 1048 QAKYGGEFLABELKAKYPEMFKVNMISTGKPIDDSVKLQKWAKEYFNGTNVLERGVGYVLS 1107
QY 1064 DOANTYFSLVSDNTFLPKSLVNPNGHTSSSTVGLVFDGKGVYVYSTSGNOAKNAPISLG 1123
Db 1108 DEATGKYFTVTKEGNFILQLT----GKEKVTGSSDGKIGITYFGTSGTQAKSAFVTFN 1163
QY 1124 NNWYFDNNGYVMTGAQ-SINGANY-FLSNGIQLRNAIY--DNGNKVLSYYGNDGRRYE 1179
Db 1164 GNTYFDFARGHMVTNSEYSPNGKDVYRFLPNGIMLSNAPYIDANGNTYL--YNSKQMYK 1221
QY 1180 NGYILF-----GQQ-----WRYFQNG-GLMAGVCLTRVHGAVOYFDASGFOAKGQFIT 1225
Db 1222 GGYTKFDVSETDKDGKESKVVYKFRYFTNEGVMAGVTVIDGFTQYFGEDGFOAKDKLV-T 1280
QY 1226 ADGKLRYPDRDRSGNOISNRFRVNRNSKGEWFLPDHNGVAVTGTVTENGQRLYFKPENGVOAKG 1285
Db 1281 FKGKTYFDDAHTGNGIKDTW-RNINGKWKYFDDANGVAATGAQVINGQKLYFNEDESGOVKG 1339
QY 1286 EFIRDANGYLRYDPNSGNSGNEVRNRFVRNSKGEWFLPDHNGIAVTGARVNG 1336
Db 1340 GVVKNADGTSYKXKBGFGELVTNEFFTTDGNVWYVYAGANGKTVTGAQVING 1390
```

Search completed: February 11, 2006, 20:56:45
Job time : 129.071 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 20:42:47 ; Search time 7.82456 Seconds
(without alignments)
2306.008 Million cell updates/sec

Title: US-10-797-821-35

Perfect score: 7143

Sequence: 1 MEKKVRFKLRKVKRWTVS.....VKVNSNTWLLIPMKFVIVM 1375

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 9704 seqs, 13122538 residues

Total number of hits satisfying chosen parameters: 97014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA_New*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	308.5	4.3	2710	7 US-11-051-453-41	Sequence 41, Appl
2	307	4.3	726	7 US-11-052-554A-217	Sequence 217, App
3	266	3.7	1155	6 US-10-793-626-1780	Sequence 1780, Ap
4	255.5	3.6	2367	7 US-11-051-453-42	Sequence 42, Appl
5	251.5	3.5	2314	7 US-11-013-759-11	Sequence 11, Appl
6	245.5	3.4	1992	7 US-11-013-759-3	Sequence 3, Appl
7	245.5	3.4	1992	7 US-11-013-759-4	Sequence 4, Appl
8	245.5	3.4	2047	7 US-11-013-759-13	Sequence 13, Appl
9	245.5	3.4	2047	7 US-11-013-759-7	Sequence 7, Appl
10	235	3.3	701	7 US-11-052-554A-231	Sequence 231, App
11	232.5	3.3	619	7 US-11-052-554A-229	Sequence 229, App
12	227	3.2	690	7 US-11-052-554A-232	Sequence 232, App
13	227	3.2	2902	7 US-11-052-554A-91	Sequence 91, Appl
14	223.5	3.1	1647	7 US-11-052-554A-260	Sequence 260, App
15	223	3.1	5291	7 US-11-052-554A-281	Sequence 281, App
16	222	3.1	2399	7 US-11-052-554A-92	Sequence 92, Appl
17	220	3.1	1588	7 US-11-052-554A-280	Sequence 280, App
18	217	3.0	396	7 US-11-022-562-228	Sequence 228, App
19	217	3.0	2053	7 US-11-013-759-9	Sequence 9, Appl
20	214	3.0	1237	7 US-11-052-554A-95	Sequence 95, Appl
21	213	3.0	5024	6 US-10-793-626-2964	Sequence 2964, Ap
22	212.5	3.0	483	6 US-10-630-203-8	Sequence 8, Appl
23	212.5	3.0	483	6 US-10-630-203-30	Sequence 30, Appl
24	212.5	3.0	512	7 US-11-102-188-5	Sequence 5, Appl
25	212	3.0	483	6 US-10-630-203-10	Sequence 10, Appl

26	212	3.0	514	7 US-11-102-188-6	Sequence 6, Appli
27	209	2.9	744	6 US-10-873-528-184	Sequence 184, App
28	208	2.9	1263	6 US-10-485-517-127	Sequence 127, App
29	207.5	2.9	693	6 US-10-873-528-185	Sequence 185, App
30	205	2.9	3194	7 US-11-052-554A-90	Sequence 90, Appli
31	200	2.8	515	6 US-10-630-203-6	Sequence 6, Appli
32	199	2.8	484	7 US-11-102-188-16	Sequence 16, Appli
33	199	2.8	542	7 US-11-102-188-9	Sequence 9, Appli
34	199	2.8	547	7 US-11-102-188-30	Sequence 30, Appli
35	199	2.8	1579	7 US-11-052-554A-9	Sequence 9, Appli
36	198.5	2.8	514	7 US-11-102-188-4	Sequence 4, Appli
37	198	2.8	515	7 US-11-102-188-3	Sequence 3, Appli
38	196.5	2.8	627	6 US-10-873-528-191	Sequence 191, App
39	196	2.7	549	7 US-11-102-188-7	Sequence 7, Appli
40	193.5	2.7	514	7 US-11-103-037-3	Sequence 3, Appli
41	193.5	2.7	515	7 US-11-195-538-4	Sequence 4, Appli
42	192	2.7	1107	6 US-10-485-517-145	Sequence 145, App
43	191.5	2.7	485	6 US-10-630-203-12	Sequence 12, Appl
44	191.5	2.7	485	7 US-11-113-775A-3	Sequence 3, Appli
45	191.5	2.7	485	7 US-11-113-799-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-11-051-453-41
; Sequence 41, Application US/11051453
; Publication No. US20050287150A1
; GENERAL INFORMATION:
; APPLICANT: AMBROSINO, DONNA
; APPLICANT: BABCOCK, GREGORY J.
; APPLICANT: BROERING, THERESA
; APPLICANT: GRAZIANO, ROBERT
; APPLICANT: HERNANDEZ, HECTOR JAVIER
; APPLICANT: LOWY, ISRAEL
; APPLICANT: MANDELL, ROBERT
; APPLICANT: MOLRINE, DEBORAH
; APPLICANT: THOMAS, JR., WILLIAM D.
; APPLICANT: ZHANG, HUI-FEN
; TITLE OF INVENTION: ANTIBODIES AGAINST CLOSTRIDIUM DIFFICILE TOXINS AND
; FILE REFERENCE: MJ1-001
; CURRENT APPLICATION NUMBER: US/11/051,453
; CURRENT FILING DATE: 2005-02-04
; PRIOR APPLICATION NUMBER: 60/542,357
; PRIOR FILING DATE: 2004-02-06
; PRIOR APPLICATION NUMBER: 60/613,854
; PRIOR FILING DATE: 2004-09-28
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 41
; LENGTH: 2710
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-11-051-453-41

Query Match 4.3%; Score 308.5; DB 7; Length 2710;
Best Local Similarity 20.4%; Fred.No. 2.5e-10;
Matches 321; Conservative 194; Mismatches 542; Indels 517; Gaps 82;
QY 92 TNSQNTTANTANFVVKPTTSE-----QAQTD-- 119
DB 1286 TNIKLDKDRNFIMPITTTTNEIRNKLSYSGDGGTYSLLSSYPSTINILSKDDLM 1345
QY 120 --NSDKIITTSKAVNRLTATGKFPANNNTAHPKTVDKIVPIKPKIKGLKQPSSLSQDD 177
DB 1346 IFNIDNEVREISIEGTTIKGKLI-----KDVLSKIDINKKL-----IIGNOT 1389
QY 178 IAAIGNVKN-----IRKVGKYYYKEDGTLOKNYALNINGKTFTFDETCALSNNTLP 230
DB 1390 IDFGSDINDKORYIFLTCELDKISLIITEINLVAKSYSLLSG-----DKNYLISNLSNT 1444

```
QY 231 SKKGNITNDNTNSAQVNOVYSTVDANFEHVDHYLTAEWSYRPKYIL---KDGKT--- 283
Db 1445 IKKINTLGLDSKNIAVNY-----TDES----NKTFGALSKTSQSIHYHKKDSKNLEIF 1495
QY 284 WTQST---EKDF-----RPLMTWPDQETQ- 307
Db 1496 YNDSTLEFNSKDFIAEDINVMKDDINTITGKYVYVDDNNTDKSIDFSISLVSKNQVKVNGL 1555
QY 308 -----QYVYMAQGIHQTNTATSPLOLNAAQTIOKIBEKITAENKTNWLKQT 359
Db 1556 YLNEVSYSYLDLVKNSDGHNTSF-----MNLFLONISF---WKLFGFENINFIADK 1606
QY 360 ISAFV-KTOSAWNS---DSEKPDFDLQKALLYNNNSKLTQASNSRYILNRTPTNQTG 415
Db 1607 YFTLCKTNLGVFEICNNKNIDIFY--GEWKTSSKSTPFSNGRNVVVEPIYNPDGTG 1664
QY 416 KK-----DPRYTADRTIGGYBFLANDVNSPVVQABQLNWLHFLMNF--NIYAN 465
Db 1665 EDISTSDFSYEPLYGIDRYIN--KVLIAPDLYTS-----LININTNYYSN 1708
QY 466 D-----PDANFDS-----IRDAVDNVDADLLQ-----IAG- 491
Db 1709 EYYPEIIVLNPNTFHKKNINLNDSSSBFYKWSSTEGSDFLVRYLBSNKKILQKIRKI 1768
QY 492 -----DYLKAAG--IKNDKAAN-----DHL--SLEAWSYNDTPYLH 526
Db 1769 LNTOSFNKMSIDFDIKLSIGYIMSNPKSPNSNELDRDLHGPKIID-----NKTYIYD 1824
QY 527 DPG---DNMINNDRLRSLLSYLAQPLNQRSGMPLNTSLNVRTDDNAETAAVPSYSF 583
Db 1825 EDSKLVKGLINNN---SLFYDPIEFNLVTGWQTI--NGKKYFYDINT-GAALTSYKI 1877
QY 584 IRA-----HSEVQDLI-----RNIIRTEINPNVGY--SFTTEEIKKAEPI 623
Db 1878 INKGHFYFNNDGVMQGVKFGPDGFEYFAPANTQNNIEGOAIVYQSKFLTLNGKKYFDP 1937
QY 624 YNKD-----LLATEKTHYNTALSYALL-----LTKNSSVP 655
Db 1938 NNSKAVTGWRIINNEKYFNPNNAAVGLQVLDNNKYFNPDPITAIISKGWTVNGSR-- 1995
QY 656 RVYVYGDMPFTDGC-QYMAHKTIN---YEAITLLKARIKYVSG-----G 694
Db 1996 ---YFED--TDTAIFNGYKTIQDKHFPDSCVVKIGVFTSNGFEYFAPANTYNNIEG 2051
QY 695 QAMRNQOV-----GNSIITSVRY--GKGALKATDTGDRTRTTSVAVIEGNN 740
Db 2052 QAIYVQSKFLTLNGKKYFDPNNSKAVTGWQITDSKKYFNTNTABEAT---GWQITDGC- 2107
QY 741 PSLRLKASDRVVVNMGAHKNOAYRPLLLTTDNGIKAVHSDQEAAGLVRYTNDRGELIFT 800
Db 2108 -----KYFNTNTABEATGWQ-----TIDGKKYFNTNTAISTGYTIINGKHFF 2153
QY 801 AADIKGYANPQVSGYLGVWVPVGAADQDVRVAASTAPSTDGKSV-HQNAAL----- 851
Db 2154 NTD-----GIMQGVKFGPGFEYFAPANTDANNIEGOAILYQNEFLTLNGKKY 2203
QY 852 ----DSRVMPGFSNFQAFATKEEY--NVVIAK-----NVDKFAEWGVTDEPMAPOVVS 901
Db 2204 FGSDSKAV---TGWRIINNNKYFNPNNAAIAIHLCTINNDKY-----YFS 2246
QY 902 STDGSLDSVIONGYAFTDRYDLG:SKPNKYGTADDLVKAIKALHSKGIKWADWVPDQM 961
Db 2247 -----YDGILONGY-----ITERNFYPDPANNESKMWTVGF-KGPNGFEYFAP--- 2289
QY 962 YALPEKEVVTRVDKYGTPVAGSQI--KNTLYVVDGKSSGKQDQQAQYGAFLLEELQAKY 1019
Db 2290 -----ANTHNNNIEGOALVYQNKFLTLNGKKYFYDNDKSAVTGWQITDGGKY 2336
QY 1020 PELFARKQISTGVMPDPSVKIKOWSA-----KYFNGTNIILGRGAGYVLKQDQANTVYFSLVS 1075
Db 2337 YFNLTNABEATG-----WQITDGGKYFYFNLTNABEATGWQITD-GKKYFYF--N 2381
QY 1076 DNTFLPK---SLAVNPNH---GTSSSVTGLVFDG-KGYVYVYSGNOA-----KNA 1118
```

```
Db 2382 TNTFIASGTYSINGKHFFYFNTDGMQGVKFGPGNGFEYFAPANTDANNIEGOAILYQNK 2441
QY 1119 FISLGNWYFNDNGYMTGAOSINGANYFSLN-GIOLRNAIYDNGKNVLSYYGNDGRR 1177
Db 2442 FLTLLNGKKYFPGSDSKAVTGLRTIDGKKYFYFNTAVAVTGWQITNGKKY--YFNTNTSI 2499
QY 1178 YENGYYLFGQWRVFO-NGIMAVGLTRVHGAVOYF-----DASGFOAKGOFITTADGKLR 1231
Db 2500 ASTGYTIISGKHFFYFNTDGMQGVKFGPDGFEYFAPANTDANNIEGOA-----IR 2550
QY 1232 YFDRDSGNQISNRFRVNRNSKGEWFLFDHNGVAVTGTVTENGORLYEKP----- 1278
Db 2551 Y-----QNRFLYLHDNIYY-PGNNSKAATGWITIDGNRYEYEPNTAMGANGYKTI 2600
QY 1279 -----NGVQAKGEFIRDANGYLRYYDPNS-----GNEVR--NRFRVNRSGWFLPD 1322
Db 2601 NKNFYFRNGLPQIGVF-KGSNGF-EYFAPANTDANNIEGOAIRYQNRFL-HLLGKIYYFG 2657
QY 1323 HNGTAVTGAUVVNG 1336
Db 2658 NNSKAVTGWQITNG 2671
```

RESULT 2

```
US-11-052-554A-217
; Sequence 217, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US 11/052,554A
; PRIOR FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 217
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Streptococcus mutans UA159
US-11-052-554A-217
```

```
Query Match 4.3%; Score 307; DB 7; Length 726;
Best Local Similarity 27.4%; Pred. No. 4,5e-11;
Matches 97; Conservative 42; Mismatches 97; Indels 118; Gaps 14;
```

```
QY 983 AGSQ-IRN-TLYVVDGK--SSGKQQAQYKYGAFLEELQAKYFELPARKQISTGVMPDPSV 1038
Db 52 ASSQKVENOTSNOVEAKTDSANKDPQKGTG-----SVATDAPSMNSA 93
QY 1039 KIKQWSAKYFNGTNIILGRGAGYVLKQDQANTVYFSLVSDNTFLPKSLVNPNGHSTSSVTCL 1098
Db 94 NNMGSQDKQ-NTVNSISSDSQQTKTDEQTD-----LPQNSFKQOS- 132
QY 1099 VFDGKGYVYVYSGNOAK---NAFISLGN-NWYFYDNNGYMVTQAQISINGANYFSLNSG 1153
Db 133 -----AHVKMTTEAEKTPSHSINTFVNDGNGWYILGADGRNVGTSGHTIGKTMVTF--- 183
QY 1154 IOLRNAIYDNGKNVLSYYGNDGRRYENGYYLFGQQWRYFQNGIMAVGLTRVHGAVOYFDA 1213
Db 184 -----AQDGK----- 188
QY 1214 SGFOAKGOFITTADGKLYBFDSDSNQOISNRFRVNRSGWFLFDHNGVAVTGTVTENGOR 1273
Db 189 ---QVKGAFAPQSDGKNHYDRDSGEMWNTNRFV-NDQGNWYVLYNNDGVPVTSITVNGOS 244
QY 1274 LYFKPENGVOAKGEFIRDANGYLRYYDPSNGNEVRNRFRVNRNSKGEWFLFDHNGIA 1327
```


QY 810 PQVSGYLGWVPVGAAADQDVRVAASTAPSTDGKSVHQNAALDSRVMFEGFSNFQAFATK 869

RESULT 6

```

US-11-013-759-3
; Sequence 3, Application US/11013759
; Publication NO. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT MORAXELLA
; TITLE OF INVENTION: RECOMBINANT MORAXELLA
; FILE REFERENCE: 1038-921MIS:3b
; CURRENT APPLICATION NUMBER: US/11/
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/36
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1992
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-3

```

Query Match 3.4%; Score 245.5; DB 7; Length 1992;

[illegible]

Db	1292	KFALSNQATGDALVKASDIVAHLNLTSGDIQTAKGASQANNSAGYVDADGNKVIYDSTDN	1351
Qy	960	QMYALPE-----KEYVTATRYDKYGTGP-----VAGSQIKNTLYVVDGSKSGKQQAQYKG	1009
Db	1352	KYYQAKNDGTVDKTKEVAKDKLVAQAQTPDGTTLAQMNVKSVINKEQVNDANKKQGINEDN	1411
Qy	1010	AFIEELQAKYPELFAKQKISTGVPMSPSVKIKQWSAKYFNGTNIILGRGAGYVLKQATYNT	1069
Db	1412	AFVKGLEKAAASD---NKTKNAAVTVGDLNAVAQTPLTTFAGDTGTGTTAKKLGSETLITIKGGQT	1468
Qy	1070	YFSLVSDNTE-----LPKSLVNPNHGTSSTSVTLGFLVDGKGYYVYVSTSGNQAKN	1117
Db	1469	DTWKLDTNNIGVAVAGTDGFTVKLAKDLTNLN---SVNAGTKIDDKGVSVFVDSGQAKAN	1525
Qy	1118	AFISLGNMYTYFDNNGYMYVT---GAQSGINGANYFYLS---NGIQLRNAIYDN---GNKV-	1167
Db	1526	TPVLSANG---LDLGGKVISNVGKGTDDTAANVQQLNEVRNLLGLGNAGNDNADGNQVN	1582
Qy	1168	---LSYYGNDGRR---YENGYYLFGQ---QWRYFQNGIMAVGLTRFVHGAVQYFDASGF	1216
Db	1583	IADIKKDPNSGSSSNRTVIKAGTVLGGKGNNDTEKLTAGGIQGVGDK-----	1629
Qy	1217	QAKGQFTTTADGKLYFRDRDSGNQISNRFVRNSKGEWFLFDHNGVAVTGVTGTFNGQRLYF	1276
Db	1630	-----DGNAGDLSNVVKTQK-----DGSKKALLATYNAGQTNYL	1666
Qy	1277	KPNGVQAKGEFIRDANGLYRYDPSNGNEV-----RNRFVRNSKGEWFLFDHNGIAVTGA	1331
Db	1667	TNPAEALDRINEQ---IRFFHVNDGNQEPVVGQGRNIDSSASGK-----HSAVIGFQ	1717
Qy	1332	RVVNGHASILSLMVFLRLESSLOSQVKVSN	1361
Db	1718	AKADGEAAV---AIGRQTOAGNQSIAIGN	1744
RESULT 7			
US-11-013-759-13			
; Sequence 13, Application US/11013759			
; Publication No. US20050249747A1			
; GENERAL INFORMATION:			
; APPLICANT: Loosmore, Sheena M.			
; APPLICANT: Sasaki, Ken			
; APPLICANT: yang, Yan Ping			
; APPLICANT: Klein, Michel H.			
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE			
; TITLE OF INVENTION: PROTEIN OF MORAXELLA			
; FILE REFERENCE: 1038-921MIS:jb			
; CURRENT APPLICATION NUMBER: US/11/013,759			
; CURRENT FILING DATE: 2004-12-16			
; PRIOR APPLICATION NUMBER: US/09/361,619			
; PRIOR FILING DATE: 1999-07-27			
; NUMBER OF SEQ ID NOS: 32			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 13			
; LENGTH: 1992			
; TYPE: PRT			
; ORGANISM: Moraxella catarrhalis			
US-11-013-759-13			

17

Db 556 GTNNSLVTAHLASYLNEVN-RTADSALQSFTVKEEDDDANAITVAKDOTTNKNAGAVSIL 614
Qy 186 NIKRVNGKYYYKEDGT-----LQKNYALNINGKTFPFDGALSNNLTSPSKGN----- 235
Db 615 KLKGKNGLTVAATKDGTVTFGLSQDSGLTIGKST-----LNNDGLTVKDTNEQIQVG 666
Qy 236 -----ITNNDNTNSFAQYNQVYSTDVANFHVHDLTAESWYRPK--YILKDGKWTQST 288
Db 667 ANGIKFTNVGNSP-----GTGIANTRIT-----RDKIGFAGSDGAV----- 704
Qy 289 EKDFRPLMTWMPDOETQOYVYNNMAQLGIHQTVN-----TATSPQLMLAAQT 338
Db 705 -----DTNKPYLDDQKLQVGNVKITNTGINAGKAITGLSPTLPISADQS 749
Qy 339 IOTKIEEKITAEBKNTNWLQRTISAFVKTSQAMNSDSEKPFDDHLOKGLLYNNNS----- 393
Db 750 -SRNIE-----LGNTIQDKKSNA-----SINDILNTGFLKNNNPIDFV 790
Qy 394 ---KLTSQANSYRILNRTPTNQTGKDPRTYADRTIGYEFLLANDVDNSPVPQAEOL 450
Db 791 STYDIDVFANGN--ATTATVTHDHTANKTSKVYVDVNVDDTTIHLTGTDDNKKLGKVTTKL 848
Qy 451 NWL-----HFLMNFNIYANDPANFDSIRVDADVNDLLOIAGDYLKAAKGIHKNDKA 506
Db 849 NKTSANGWTATNF--NVNSDEDA-----LVNAKD-----IAENLNTLAKEIHTTKGT 894
Qy 507 ANDHLSILEAMSYNDTPYLDHDDGNNIMNDNRLRLSLLSYLAQPLNQRSGMNPFI---TN 563
Db 895 ADTALQTFVKKVDENN--NADDANAITVGOK-----NANNQVNTLILKGEN 939
Qy 564 SLVNETDONAE--TAAPVPSYFTRAHDSEVD--LIRNIIRTE-----INPN 606
Db 940 GLNIKTGNGTGTFTGINTTSLGKAGKSTLNDGGLSIKNPQTSSEQIQVGADGVKFAKAVNN 999
Qy 607 VV---GYSFTTEEIKKAEIYNKOLLATEKKYTHYNTALSYALLTNTKSSVPVRYGDMF 663
Db 1000 GVWGAIGDGTTR-----ITRDEIGFTNGSLD-----KSKPHL----- 1033
Qy 664 TDDQYMAHKITNEYBAITELKARIKYVSGGQAMRNQOVGSEIITSVRYGKALKATDT 723
Db 1034 SKDGINAGGKKTINTQSGEIAQNSHDAVTGGKIYDLKTELENKISSTAKTAQNSLHEFSV 1093
Qy 724 GDR-----TTRTSQVAVIEGNNPSLRKASDRVYVNMGAHKNQAYRPLL 768
Db 1094 ADEQGNFTVSNPYSSYTSKTSVDITEAGEN-GITTKV-NKGVVVRVGDQTKGLTTPKL 1151
Qy 769 LTTDNGIKAYHSDOE-----AAGLVRYTNDRGELIFTAADIKGVANPOVSGYLGW 819
Db 1152 TVGNNGGVIDSQNGQNTIGLSNTLANVTNDRKGSVRTTEQG----- 1195
Qy 820 VPVGAADQDVRVAASTAPSTDGKSVHQAALDSRVMPFEGFSNFQAFATKKEEYTNVIA 879
Db 1196 ---NIKDEDKTRASIYD-----VLSAGP-NLOG-----NGEAVDFVST 1231
Qy 880 KNVDKFAEWGVTDFEMAPQYVSSSTDGSLDSVIONGYAFTDRYDLGI-----SKPN 930
Db 1232 YDTVNFADGNATTAKVTYDDTSKTSKVYVDVNVDDTTIEVKDKLGKVTTLTSTGTGAN 1291
Qy 931 KY-----GTADDLYKA--IKA-----LHSGIKGIKWADWVPD 959
Db 1292 KPALSNQATGALVKASDIVAHLNLTSGDIQTAKGASQANNSAGYVDADGNKVIYDSTDN 1351
Qy 960 QMYALPE-----KEVVTATRVDKYGP--VAGSQIKNTLYYVDGSKSGKQQAQYGG 1009
Db 1352 KYQAQKNDGTVDKTEVAKDKLVAQAQTPDGTILAQMNVKSVINKEQVNDANKQOINEDN 1411
Qy 1010 AFLELOAKYPELFARKQISTGVPMDPSPVKIKQMSAKYFNGNTIILGRGAGVYLKDAQNT 1069
Db 1412 AFVKGLEKAAAD--NKTKNAAVTVGDLNVAQAQTEPLTFAGDTGTTAKKLGTELTIKGGQT 1468
Qy 1070 YFSLVSDNTF-----LPKSLVNPNHGTSSSVTGLVFDGKGVYVYSTSGNQAKN 1117
Db 1469 DTNKLTDNNIGVAGTGDGTVKLAKDLTNLN---SVNAGGTKIDDKGVSFVDSQAKAN 1525

RESULT 8

US-11-013-759-4
; Sequence 4, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loesmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN OF MORAXELLA
; FILE REFERENCE: 1038-921MIS:jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2047
; TYPE: PRP
; ORGANISM: Moraxella catarrhalis
US-11-013-759-4

Query Match 3.4%; Score 245.5; DB 7; Length 2047;
Best Local Similarity 19.9%; Pred. No. 1e-06;
Matches 304; Conservative 190; Mismatches 615; Indels 421; Gaps 66;

Qy 10 RKVKRWTVTSIASAVVLTLSLGSGLVKADSTDDRQQAVTESQASLVTTSBAKETLTAT 69
Db 513 KQLKVGSAITIDNGIDAGNKKISNLAKGSAND---AVTIEQL-----KAAKPTLNA- 562
Qy 70 DTSTATSATOPTATVTDNVSTTNSQNTTNTANTANFVVKPTTTSQAKTDNSDKIITTSK 129
Db 563 --GAGISVT--PT-----EISVDAKSGNVTAFTYNIQV---TTLENSDGTSDKESVKS 610
Qy 130 AVNRLTATGKVPANNNTAHPTVTDKI--VPKPKIGKLKOPSLSDQDIAALG--NVK.185
Db 611 GTNNSLVTAHLASYLNEVN-RTADSALQSFTVKEEDDDANAITVAKDOTTNKNAGAVSIL 669
Qy 186 NIKRVNGKYYYKEDGT-----LQKNYALNINGKTFPFDGALSNNLTSPSKGN----- 235
Db 670 KLKGKNGLTVAATKDGTVTFGLSQDSGLTIGKST-----LNNDGLTVKDTNEQIQVG 721
Qy 236 -----ITNNDNTNSFAQYNQVYSTDVANFHVHDLTAESWYRPK--YILKDGKWTQST 288
Db 722 ANGIKFTNVGNSP-----GTGIANTRIT-----RDKIGFAGSDGAV----- 759
Qy 289 EKDFRPLMTWMPDOETQOYVYNNMAQLGIHQTVN-----TATSPQLMLAAQT 338
Db 760 -----DTNKPYLDDQKLQVGNVKITNTGINAGKAITGLSPTLPISADQS 804

Qy 339 IOTKIEEKITAEKNTNWLROTISAPVKTQSAWNSDSEKPFDDHLQKQALLYSNNS----- 393
Db 805 -SRNIE-----LGNTIQDKXKNA-----SINDILNTGNLKNNNPIDFV 845
Qy 394 ---KLTSAANSYRILNKTPTNQTGKDPRTADRTIGGYEFLLANDVNSPNVVAEQOL 450
Db 846 STYDIVDFANG--ATTATVTHDTANKTSKVYDVNVDTTIHLTGTDNKKLGKVTYKL 903
Qy 451 NWL-----HFLMNFNIYANDPANPDSIRVDANVDADLLQIAGDYILKAAGIKHNKDKA 506
Db 904 NKTSANGNTATNF-NVNSSDEDA-----LVNAKO-----IAENLNTLAKEIHTTKGT 949
Qy 507 ANDHLSILEAWSYNDTPYLDHGDGMNMDNRRLSLLSYSLAKPLNQSRGWNPLI---TN 563
Db 950 ADTALQOTFTVKVVDENN--NADDANAITVGOK-----NANNQVNTLTKEEN 994
Qy 564 SLVNRDTONAE--TAAPVPSYFIRAHSDVQD---LIRNIIRTE-----INPN 606
Db 995 GLNKTDRKNGTVTFGINTTSGLKAGKSTLNDGGLSIKNTPTSEQIQVGADGVKFAKVANN 1054
Qy 607 VV---GYSFTTEEIKKAFIYNKDLLATEKKYTHYNTALSYALLLTNKSXSPRVVYGDWF 663
Db 1055 GVWGAIDGTR-----ITRDEIGFTGTNGSLD-----KSKPHL----- 1088
Qy 664 TDDGOYMAHKTINYEAIETLKARIKYVSGQAMRNQOVGNSBITTSVRYKGGALKATDT 723
Db 1089 SKDGINAGGKITTQSGEIAQNSHDAVTGGKIYDLKTELENKISSTAKTAQNSLHEFSV 1148
Qy 724 GDR-----TTTSGVAVIEGNNPSLRKASDRVVVNMGAHKNQAYRPLL 768
Db 1149 ADEQNNFTVNPVSSYDTSKTSVDITFAGEN-GITTKV-NKGVVRVVDIGDTKGLTTPKL 1206
Qy 769 LTTONGIKAYHSDQ-----AAGLVRYTNRDELIFTAADIKGYANPQVSGYLGW 819
Db 1207 TVGNNGKVIDSQNGQNTITGLNLTANVTNDKGSVRTTEQ----- 1250
Qy 820 VPVGAADQDVRVAASTAPSDGKSVHQAALDSRVMEGFSNFOATKKEEYNNVIA 879
Db 1251 ---NIKEDKTRAASIVD-----VLSAGF-NLQG-----NGEAVDFVST 1286
Qy 880 KNVDKFAEWGVTDPFEMAPQYVSSDTGSLDSVQNGYAFTRDYDLGI-----SKPN 930
Db 1287 YDTVNFADGNATTAKVYDVTDSKTSKVYDVNVDTTIEVKDKLGKVTTLTSTGTGAN 1346
Qy 931 KY-----GTADDLVKA--IKA-----LHSGKIKVMADWVPD 959
Db 1347 KFALSNQATGALVKASDIVAHLNTLSGDIQTAKGASQANNSAGYVDADGNKVIYDSTDN 1406
Qy 960 QMYALPE-----KEVVTATRVDKYGT-----VAGSQIKNTLYVVDGKSSGKQOAKYGG 1009
Db 1407 KYQOAKNDGTVDKTEVAKDKLVAQAQTPDGTQLAQMNVKSVINKEQVNDANKQGINEDN 1466
Qy 1010 AFLELOAKYPELPARKQISTGVPMDPVSKIKOWSAKYFNGTNIILGRGAGYVLKQOATNT 1069
Db 1467 AFVKGLEKASD---NKTKNAAVTVDNLNAVAQTPLTFAGDTGTGTAKKLGTELTIKGGQT 1523
Qy 1070 YFSLVSDNTF-----LPKSLVNPNGHTSSSVTGLVFDGKGVYVYSTSQAKN 1117
Db 1524 DTNKLTDNNIGVAVAGTDGFTVKLAKDLTNLN---SVNAGGTKIDDKGVSFVDSQOAKAN 1580
Qy 1118 AFISLNNWYFDPNNGYMT-----GAQSGINGANYFFLS---NGIOLRNAIYDN--GNKV- 1167
Db 1581 TPVLISANG---IDLGGKVISNVGKGTQDTRAAVQQLNEVRNLLGLGNAGNDNADGNQVN 1637
Qy 1168 ---LSYYGNDGR-----YENGYILFGQ---QWRYFQNGIMAVGLTRVHGAQYEDASGF 1216
Db 1638 IADIKKDPNSGSSNRTVIKAGTVJGGKGNNDTEKLATGGIQGVGDK----- 1684
Qy 1217 QAKGQFITTAGDKLRYFRDSDGNQISNRFVRNSKGWFLFDHNGVAVTGTFTFNGQRLYF 1276
Db 1685 -----DGNANGDLSNVWVKTK-----DGSKKALLATYNAAGQTYNL 1721

Qy 1277 KPNGVQAKGEFIRDANGYLRYVYDPSNGNEV-----RNRFRVNSKGWFLFDHNGIATVGA 1331
Db 1722 TNNPAAIDRINEOG---IRFFHVNDGQBPVQGRNGIDSSAGK-----HSAIGFQ 1772
Qy 1332 RVVNGHASILSLMVFRLRESSLSQSVKVVSN 1361
Db 1773 AKADGEAAV---AIGRQTOAGNQSIATGDN 1799
RESULT 9
US-11-013-759-7
; Sequence 7, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN OF MORAXELLA
; FILE REFERENCE: 1038-921MIS:JB
; CURRENT APPLICATION NUMBER: US/11/013.759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361.619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 7
; LENGTH: 2047
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-7

Query Match 3.4%; Score 245.5; DB 7; Length 2047;
Best Local Similarity 19.9%; Pred. No. 1e-06;
Matches 304; Conservative 190; Mismatches 615; Indels 421; Gaps 66;

Qy 10 RKVKRQWTVSIAAVVTLTSLGSLVKADSTDDROQAQVTSOASLVTTSAAKETLTAT 69
Db 513 KQLKGVSAITIDGIDAGNKKISNLAKGSSAND---AVTIEQL-----KAAKPTLNA- 562
Qy 70 DTSTATSATSOTATVTDNVTNQTSTNTTANTANFVVKPTTTSBOAKTDSNDKIITTSK 129
Db 563 --GAGISVT--PT-----EISVDKSGNVTAPTNYGVK---TTELSNDSGTSDFSVKGS 610
Qy 130 AVNRLTATGKFPANNTAHPKTVTDKI--VPIKPKIGKLPSPSLSQDDIAALG--NVK 185
Db 611 GTNNSLVTAEHLASVINEVN-RTADSALQSFTVKEEDDDANAITVAKDTTKNAGAVSIL 669
Qy 186 NIKVNGKYVYKEDGT-----LQKYNALNKGKTFPFDGTGALSNTLPSKKN----- 235
Db 670 KLKGNGLTVATKDKGTVTFLGSLQDSGLTIKST-----LNNDGLTVKDTNEIQVG 721
Qy 236 -----ITNNDNTNSPAQVNVYSTVDANFEHVDHYLTAEBSWYRPK--YILKDKGTWTQST 288
Db 722 ANGKFTNVGNSNP-----GTGIANTRIT-----RDKIGFAGSDGAV----- 759
Qy 289 EKDFRPLMTWMPDOETQRYVYVNMNAQLGIHQTYN-----TATSPQLNLAQT 338
Db 760 -----DTNKPVLDDQLQVGNVKITNTGINAGGKAITGLSPTLPSIADQS 804
Qy 339 IOTKIEEKITAEKNTNWLROTISAPVKTQSAWNSDSEKPFDDHLQKQALLYSNNS----- 393
Db 805 -SRNIE-----LGNTIQDKKXNA-----SINDILNTGNLKNNNPIDFV 845
Qy 394 ---KLTSAANSYRILNKTPTNQTGKDPRTADRTIGGYEFLLANDVNSPNVVAEQOL 450
Db 846 STYDIVDFANG--ATTATVTHDTANKTSKVYDVNVDTTIHLTGTDNKKLGKVTYKL 903
Qy 451 NWL-----HFLMNFNIYANDPANPDSIRVDANVDADLLQIAGDYILKAAGIKHNKDKA 506
Db 904 NKTSANGNTATNF-NVNSSDEDA-----LVNAKO-----IAENLNTLAKEIHTTKGT 949

Db 605 -----WLQNN-GSWYLYNANGDMATGWLQYNGSWYLYNANG 639
Qy 1281 VOAKGEFIRD-----ANGYLR-----YDPSNG-----NEVRNPFVNSKGE 1317
Db 640 DMATG-WVKDGTWYLEASGAMKASQWPKVSDRWYVNGSGALAVNTTVDGYGVNANGE 698
Qy 1318 W 1318
Db 699 W 699
RESULT 11
US-11-052-554A-229
; Sequence 229, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 229
; LENGTH: 619
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae R6
US-11-052-554A-229

Query Match 3.3%; Score 232.5; DB 7; Length 619;
Best Local Similarity 28.2%; Pred. No. 1.1e-06;
Matches 68; Conservative 35; Mismatches 99; Indels 39; Gaps 10;
Qy 1096 TGLVFDGKGYVYSTSGNOAKNAFISLGNWYTFDNGWMTGAOSINGANYFISNGIQ 1155
Db 402 TGWQENGWYFYNTDGSNA-TGWLQNNGSWYLYNSNGAMATGWLQYNGSWYLYNANGAM 460
Qy 1156 LRNAIMDNKNVLSYYGNDGRRYENGYLFGQWRFYQ-NGTMVGLTRVHGAVOYFDAS 1214
Db 461 ATGWAKNGS-----WYLYNANGAMATGWLQYNGSWYLYNANGAMATGWAKYNGSWYLYNAN 517
Qy 1215 GFQAKGQFTTADGKLRYFDRDSGQISNRFVRNSKGWFLFDHNGVAVTGTVPNGQRL 1274
Db 518 G-----AMATGWLQY-----NGSWYLYNANGAMATGWAKYNGSWY 552
Qy 1275 YPKPNGVQAKGFIRDANGYLRYPDPSNGNEVRNPFVNSKGEWFLFDHNGIANTGARVV 1334
Db 553 YLNANGAMATG-WVKDGTW--YYLEASGAMKASQWPKVSDRWYVNGSGALAVNTTVDGYGVNANGE 603
Qy 1335 N 1335
Db 604 N 604
RESULT 12
US-11-052-554A-232
; Sequence 232, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004

; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 232
; LENGTH: 690
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae R6
US-11-052-554A-232
Query Match 3.2%; Score 227; DB 7; Length 690;
Best Local Similarity 19.7%; Pred. No. 2.7e-06;
Matches 146; Conservative 110; Mismatches 288; Indels 198; Gaps 31;
Qy 702 VCNSEIITSVRYGKALKATDGTTRTSGVAVTEGNNPSLRKLASDRVVVNMGAHKN 761
Db 20 VPNEPIALAYVNEPILADTPSSSEIKETKGSIIQQNNIKYKVLV---VEGNIGTVQVG 76
Qy 762 QAYRPLLLTTDNGIKAYHSDQEAAGLVRYTN-DRGELIFTAADIKGYA---NPQVSGVILG 817
Db 77 NGVTPV-----EFGAGQDGRKPTTPTKITVGDKYFTTVEVASQAFSYYPDETGR- 126
Qy 818 VWPVVGAAADQDVRVAASTAPST----DGKSVHQAALDSRVFMFGFSNFQAFATKBEY 873
Db 127 VVYP-----SSITPSSIKKIQKGFHGSKA--KTIIFDKGSQLEK----- 165
Qy 874 TNVVIKNDVKAENGVTDFEMAPQYVSVSTDSGSLDSVLQNGYATFDRYDL-----GI 926
Db 166 ----FEDRAFDFSELEETELPASLEJYIGTSAFSFSQKLKLTFFSSSSKLELISHAFANL 221
Qy 927 SKPKNYGTADDLVKAIKALHSK-----GI-----KVMADWV 957
Db 222 SNLEKL----TLPKSVKTLGSNLFLLTSLKHVDVEEGNESFASVDGVLFSKDKTQLIYY 277
Qy 958 P-----DOMYALPE-----KEVVTATRVDKYGTTPVAGSOIK-----NT 990
Db 278 PSQKNDESYPKPKETKELASYSFNKNSYLLKLELNEGLEKIGTFAPADAIAKLEELSPNS 337
Qy 991 LYVVDKSSGKDOQAKYGGAFLEELQAKYPELFARKQISTGVPMDFSVKIKOMSAKY--F 1048
Db 338 LETI-----ERLAFYGNLELKEILPDNKNFQKHWVNGLP-----KFLTL 378
Qy 1049 NGTNILRGAGYV-----LKD---QATNTVFSLVSDNTFLPKS----- 1083
Db 379 SGNNINSLPSPFLSGVLDLSLKEIHKKSTEFSSVKDTFAIPETVYFVTSSEHIKDLKS 438
Qy 1084 -----LVNPHGTSSVTGLVFDGKGYVYSTSGNOAKNAFISL 1122
Db 439 NLSTSDNDIIVKVDNIKQETDVAKPKNQGVGVGWVKD-KGLWYLYNESGSMATGWVKD 497
Qy 1123 GNNWYFPDNGWYMTGAOSINGANYFISNGIOLRNLAYDNGKNVLSYYGNDGRRYENG 1182
Db 498 KGLWYLYNESGSMATGWVKDKGLWYLYNESGSMATGWVKDKG---LWYLYNESGSMATGW 554
Qy 1183 YLFGQWRFYF-QNGIMAVGLTRVHGAVOYFDASGFQAKGQFTTADGKLRYFDRDSGNOI 1241
Db 555 VKDKGLWYLYNESGSMATGWVKDKGLWYLYNESGSMATG---WVKDKGLWYLYNESGS-M 610
Qy 1242 SNRFVRNSKGWFLFDHNGVAVTGTVPNGQRLYFKPNGVQAKGFIRDANGYLYYDDN 1301
Db 611 ATGWVKD-KGLWYLYNESGSMATGWVKDKGLWYLYNESGSMATG-WVKVSGKW--YYTVN 666
Qy 1302 SGNEVRNR-----FVRNSKGEW 1318
Db 667 SGDLLVNTTTPDGYVRNANGEW 688
RESULT 13
US-11-052-554A-91
; Sequence 91, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE

; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL

; FILE REFERENCE: #0853/40359A

; CURRENT APPLICATION NUMBER: US/11/052,554A

; PRIOR FILING DATE: 2005-02-07

; PRIOR APPLICATION NUMBER: US 60/589,227

; PRIOR FILING DATE: 2004-07-20

; PRIOR APPLICATION NUMBER: IN 173/DEL/2004

; PRIOR FILING DATE: 2004-02-06

; NUMBER OF SEQ ID NOS: 763

; SOFTWARE: Patent in version 3.3

; SEQ ID NO 91

; LENGTH: 2902

; TYPE: PRT

; ORGANISM: Helicobacter pylori J99

US-11-052-554A-91

Query Match 3.2%; Score 227; DB 7; Length 2902;

Best Local Similarity 18.2%; Pred. No. 2.2e-05;

Matches 313; Conservative 221; Mismatches 591; Indels 598; Gaps 82;

QY 13 KKEWTVSIAAYVTLTSLGSLVKADSTDDROQAVTESQASLVTTSEAAKETLTATDTS 72

Db 217 KTSQVNVGANSVITINSVS---LNGDTCSLARVGVCANSTSGPSYFKGTTNATNT 273

QY 73 TATSA---TSQPAT-----VTDNVSTTQSTNTTANTA-----NFVVKPT----- 110

Db 274 FNSSGSFTFEENATPSGAKLNGGAFTFNKFNATNTNATFNSGFTFKGTSSPFGANFSN 333

QY 111 ---TTSEQAKTNSD-----TKI-----VPIKPK-----IGK--LK 168

Db 334 ASYTFNQATFQNSPFGGFTTFNDQTNOSTQHPQIONSFSFGSATTLKGATFEQAFNN 393

QY 147 TAPKTV-----TDKI-----VPIKPK-----IGK--LK 168

Db 394 SNHQLTIQASFNATFNNTGKTIKEDASFNNTSFNTPDVNTNMITSGGVTLSGKNDLK 453

QY 169 QESL-----SQDIAALGNVNRKNGKYYKYKEDGTQLQKNYALN--INGK 214

Db 454 NGATLDFGSSKIPTLQGTTFNLTSLGSEKSVTILNSR-----GGITYNHLNLHAINSL 506

QY 215 T-----PPDETGALSNTLPSKKNITNDNTNSFAGYNQV 251

Db 507 TNAKTNESKPKQSPAQLMDMITYNGVTGQLLNEAATSKP--TDSSPSKSTNSTQV 564

QY 252 YSTDV-----ANFEH-----VDHLYTAE--SWYR 273

Db 565 YQVGYKIGDTIYKLOTFPSHNSIIQALBSGYTPPPVINGSKFDLSASNYINADPWYN 624

QY 274 PKYILKDGKWTQSTEKDPRLLMTWMPDQETQRYVYVYNAQLGHQTYNTATSPLOLN 333

Db 625 HKYYPKSNQFTESGTY-YLPSVQIWG-----SVTNS---PKQTFASASNLVIG 670

QY 334 LAAQTLOTIEKITAETKTNLRLQRTISAFVKTQSAWNSDSEKPPDHDHLOKQALLYSNNS 393

Db 671 Y-----NATWTDHNVSS--SDTVAFGDTSGSALNGHC--GPWPYYQCT 709

QY 394 KLTQANSNVRI---LNRPTNQTGKDPRTYADRTIGGYEFLANDVDNSN---PVQV 446

Db 710 GTTNGTYSAYHVITANLSRGRIG-----TGAANLIFNGVDISINIANATITQ 758

QY 447 -----AEQLNLHF---LMNFGNIYANDP-----DANFDSI 474

Db 759 HNAGAYSSMTFTSQNMDSNLGLNSGKLLVYGTTFNQAKDGKFIENAGQATFEN- 817

QY 475 RVDADVNDADLLOIAGDYKKAAGIH-----KNDKAANDHLSILEAWSNDTPY 524

Db 818 -----TNFNGGYSQFSGSLNFSNNQNFNSGSGFEIGAKNTIFNNAFNNTSTSFENNSSA 872

QY 525 LHDGDGNMLNMRRLSLLSYLAKPLNQSRGNWPLITNSLVNRTDDNAETAAPVSYSFI 584

Db 873 TTSFVGDFTNANSNLOIA-----GNAVFGNS-----TNGSQNTANFNNTGVS 914

QY 585 RAHDSEVDQLIRNII-----RTEINPNVVGYSFTTEIKKAFIYNKOLLATEKKYTHYN 639

RESULT 14

US-11-052-554A-260

; Sequence 260, Application US/11052554A

; Publication No. US2005028866A1

; GENERAL INFORMATION:

; APPLICANT: Sachdeva, et al.

; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE

; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL

; FILE REFERENCE: 30853/40359A


```
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 260
; LENGTH: 1647
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes MGAS8232
US-11-052-554A-260

Query Match      3.1%; Score 223.5; DB 7; Length 1647;
Best Local Similarity 18.4%; Pred. No. 1.6e-05;
Matches 312; Conservative 228; Mismatches 597; Indels 563; Gaps 82;

QY 1 MEKKVRFKLRKVKRWVTVYSIASAVVLTLSGSLVKADSTDDROQAV-----TESQASIV 56
DB 1 MEKKQRFSLURKYSKGTFSVLGSLVFLMMT-----TTVADELSTMTSEPTITNHTQQAQHL 56
QY 57 TTSEAAKETLTATDTSATSATS-----QPTATVTIONVSTNQTNTT--ANTANFVVKPT 110
DB 57 TWTLSAESKSDTSQITPKTNREKEQOGLVSEPTTELADTDAAPMANTGPDATQKS 116
QY 111 TTSEQAKTNSKIITTSKAVNRLFATGKFPVANNHTAHPKTVTDKIPIKPKIKLQKP 170
DB 117 ASLPPVNTDVHVKWTGAWDKGYKGQKVVAVIDTGIDPAHQSMRISDV--STAKVK-- 172
QY 171 SLSQDDIAALGNVKNIRKNGKYYYKEDGTGLQKNYALNNGTKTFFPDDETGALESNNITLP 230
DB 173 ---SKEDMLAROKAAGI-----NYGWINDKVVF----- 198
QY 231 SKKGKITNNDN--TNSFAQYNOVYSTDVANFEHVDHLYLTAESWYRPKYILKDKGTWTQST 288
DB 199 -AHNVVENSNDIKENQFEDFE---DWEHFE-----PDAEA--EPKAIKHKHIIYRPGST 246
QY 289 EKDFRPLMTW-----WPDQETQOYVNMNAQLGIHOT-----YNTATSPLOQ 331
DB 247 QAPKETVIKTEETDGHSDIDMTQTDHDKYKESH---GMHVTGIVAGNSKEAAATGERF 301
QY 332 LNLA-----AQIOTK--IEKITAENKTNLWLRQTISAFVKTOSAWNS 372
DB 302 LGIAPEAQWFMVRFANDVNGSAESLFIKAIEDAVALGAD-----VINLSIGTANGAQL 355
QY 373 DSEKPFDDHLQK-----GALLYSNNSKLTSQ-----ANSNYRIL-----NRTPTN-- 412
DB 356 SGSKPLMEALBKAKKAGVSVVVAAGNERVYGSDDHDDPLAINPDYGLVGSPTSRTPTSA 415
QY 413 -----QTGK-----KDP-----RYTADRTIG 428
DB 416 AINSKWVIQRLMTVKELNADLNHGKATYSESDVPKNIKDSIGDYDKSHQFAYVKESTDA 475
QY 429 GYEFLLANDVNSNPVQAEQLNWLHPLMNFN-----IYANDPDANFDSIRDAVD 480
DB 476 GYK---AQVDKQKIALIEDPNTKTYDEMIALAKKHGALGVLIFFNNKPGQSNRSMRLTA-N 531
QY 481 NYVDADLLQIAGDYKAAKGHKNDAANDHLSILEA-----WSYNDTPYLH 526
DB 532 GMGIPSAFISHEFGKAMSQLNGNGTSLSEFDSVVSKAPSKQKGNEMHFNWGLTSDGYLK 591
QY 527 DP---GDMNIM--DNRLRLSLLYSLAKPLNQRGMNPLITNSLVNRTDDNAETAAPVS 580
DB 592 PDITAPGGDIYTYNDNHYGSGTGTSMASP--QIAGAS--LLVKQYLEKTQPN-----LPK 643
QY 581 YSFIRAHDSVDLLRNII---RTEINPVVGYSTTTEIEIKKAFEIYNKOLLATEKKYT 636
DB 644 -----EKIADIWNKLLMSNAQIHVNPET---KTTPSPROQGAGLLNID----- 683
QY 637 HYNITALSYALLTNKSSVPRVYVG---DMFTDDGGYMAH-----KTINYEAE-----IE 681
DB 684 ---GAVTSGLYVTGKNYSGISIGNITDWTFTD--VTVINLSNKDKTLRYDTLLTDHVD 738
```

RESULT 15

```
US-11-052-554A-281
; Sequence 281, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sacchrdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
```

```
QY 682 -----TLLKARIKYVSGQAMRNQOVNSBIITSVRYKKGALKAT-----DTGDRTRPT 730
DB 739 PQKGRFTLTSLRSLKTYOGGE-----VTVPANGKVTVRVTVMDVSOQFTKSLTLTKM 786
QY 731 SGVAVIEG-----NNPSL-----RLKASDRV----- 751
DB 787 SNGYILEGFVFRDSQDDQLNRVNIPIFVGPKQGFENLAVAEESIVRLSKSGKTYGFYFDES 846
QY 752 -----VVMNGAAHKQAYRPLLLTTDNGIKAYHSDQEAAG--LVRYTNDRGE 796
DB 847 GPKDDIYVGKHTGLVTLG--SETNVSTKTI---SDNGLHTLGTFRKADGKFILEKNAQGN 902
QY 797 LIFTRAA-----DIKGYANFQVSGYLGWVPVGAADQDVRVAASTAPSTDCKSVHQA 849
DB 903 PVLAIISPNGDNNQDPAFAFGVFLRYQGLKASVHASDKE-----HKNP 946
QY 850 ALDSRVMEFGSFQAFATKKEEYTNVVIKXNV-----DKFAEWGVTDFEMAPQYVVSSTDG 905
DB 947 LWSPESEFKGNFN-----SDIRFAKSTLLGTAFSGKSLTGAEPLPGYHYHYVS 997
QY 906 SFLDSVIOQGYAFTDRYDLGISK--NKYGTADDLVKAIK--ALHSKGIKVMADWPDQMY 962
DB 998 YYPDVVGAQRQEMTFDMLDRQKPVLSQATFDPETNRFPKPLKDRG---LAGVRKDSVF 1054
QY 963 ALPEKE--VVTATRVYDKYGTVPAGSOIKNTLYV--VDGKSSGKQOQAKYGGAFLESLQAKY 1019
DB 1055 YLERKDNKPYTVTINDSY-----KYVSVEDNKTFFVERQA---DGSFI----- 1093
QY 1020 PELFARKQISTGVPMPSVKIKQWSAKYFNGTNIILGRGAGYVLKQATNTVPSL-----V 1074
DB 1094 -----LPJD--KAKLGD---YMWVEDFAGNVAIAKLGDLHPQLTGLKTIPLK 1137
QY 1075 SDNTFLPKSLVNPNGHTSSSVTGLV-----FDGKGYVYVYSTGNGQ 1114
DB 1138 TDGNYQTKETLKDNLMTQSDTGLVTNQALAVVHRNQPSQLTKMNQDFFISPNEDGNK 1197
QY 1115 AKNAFISIGNWY-----YFDNNGYVMT-----GAQSGINGANYFISNGIQLRN 1158
DB 1198 DFVAFKGLKNVYNDLTVNVYAKDDHQKQTPWSSQAGASASAIESTAWY---GITARG 1253
QY 1159 AIYDNGN--KVLSSYVNDGRRYENGYYLFGQWRYFQNGIMAVGLTRVHGAVOYFQDASGF 1216
DB 1254 SKWPGDYQVYVYRDEHGEKQKQYTTISVD---KKPMITQGRFDITNGVDHTPTDKT 1309
QY 1217 QAKGO-----PITTAGKLYRFDSDGNQ---ISNR---FVRNSKGEWFLFDHNGV 1261
DB 1310 KALGSSGIVREEVFLAKKNG--RKFDVTEGKDGITVSDNKNYIPKPNPDGSGYTIKRDGV 1367
QY 1262 AVT-----GTVTFNGQRLYFKPNGVOAKGE-----F 1287
DB 1368 TLDYVYLLVEDRAGNVSFATLR-----DLKAVGKDKAVVNFGLDLPVPEDKQIVNFTYL 1421
QY 1288 IRDANG---YLRVYDPNSGNEVRNRFVRNSKGEWFLFDHNGIAVTGARVNGHASILSL 1343
DB 1422 VRDADCKPIENLEYN--NSGNSL-----ILPYKQYTV-----EL 1454
QY 1344 MVFRLESSLOSQVSVSNMT 1363
DB 1455 LTYDTNAAKLESCKIVSFTL 1474
```

RESULT 15

```
US-11-052-554A-281
; Sequence 281, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sacchrdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 19:15:50 ; Search time 28.3806 Seconds
(without alignments)
4661.567 Million cell updates/sec

Title: US-10-797-821-35
Perfect score: 7143
Sequence: 1 MEKKVRFKLKVKRWTVTS.....VKVNSNTMILIPMKFVIM 1375

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7143	100.0	1375	2 J0345	dextranucrase [EC
2	5285	74.0	1475	2 B33135	gtfB protein precu
3	4168.5	58.4	1592	2 A38175	glucosyltransferas
4	3591.5	50.3	1431	2 A5866	dextranucrase [EC
5	3216	45.0	1577	2 T30858	glucosyltransferas
6	3159	44.2	1449	2 T30857	glucosyltransferas
7	3146	44.0	1449	2 T30552	glucosyltransferas
8	3070	43.0	1518	2 A44811	glucosyltransferas
9	3044	42.6	1508	2 T31098	probable dextran
10	2899	40.6	1599	2 S22737	glucosyltransferas
11	2893.5	40.5	1365	2 A41483	glucosyltransferas
12	2870.5	40.2	1290	2 JCS473	dextranucrase [EC
13	428.5	6.0	2817	2 B97033	uncharacterized pr
14	308.5	4.3	2710	2 A37052	toxin A - Clostrid
15	292.5	4.1	2364	2 I40884	cytotoxin L - Clos
16	283	4.0	2178	2 S55805	alpha-toxin - Clos
17	274.5	3.8	2366	2 S10317	glucan-binding pro
18	268.5	3.8	2366	2 S10317	toxin B - Clostrid
19	267.5	3.7	1609	2 S25345	probable membrane
20	255.5	3.6	2367	2 S70172	toxin B - Clostrid
21	242	3.4	648	2 S10869	enterotoxin A - Cl
22	235	3.3	701	2 H98120	choline binding pr
23	235	3.3	1463	2 T30290	AAS surface protei
24	233.5	3.3	1385	2 D89824	hypothetical prote
25	233.5	3.3	1939	2 D97316	probable S-layer p
26	233	3.3	1335	2 T30211	autolysin E - Stap
27	232.5	3.3	619	2 A37887	surface protein ps
28	232.5	3.3	619	2 A41971	surface protein ps
29	230	3.2	1315	2 T28679	fibrinogen-binding

30	229.5	3.2	5188	2 B85547	probable RTX famil
31	227	3.2	690	2 F98114	choline-binding pr
32	227	3.2	2902	2 C71953	toxin-like outer m
33	226	3.2	2529	2 B64635	toxin-like outer m
34	226	3.2	6713	2 B89921	hypothetical prote
35	225.5	3.2	2334	2 S39220	cell wall-associat
36	223.5	3.1	4919	2 T31105	hypothetical prote
37	223	3.1	5291	2 F90696	hypothetical prote
38	222	3.1	1582	2 A71153	adhesin homolog lm
39	222	3.1	2399	2 H71879	toxin-like outer m
40	220	3.1	1588	2 A86036	probable adhesin Z
41	220	3.1	1588	2 H91188	probable adhesin E
42	219.5	3.1	1819	2 D97033	uncharacterized pr
43	218	3.1	1166	2 T28680	fibrinogen-binding
44	217	3.0	3890	2 C89921	hypothetical prote
45	216	3.0	4152	2 T31102	filamentous hemagg

ALIGNMENTS

RESULT 1

J0345
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)
N:Alternate names: sucrose 6-glucosyltransferase
C:Species: Streptococcus mutans
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 09-Jul-2004
C:Accession: J0345; C33135
R:Ueda, S.; Shiroza, T.; Kuramitsu, H.K.
Gene 69, 101-109, 1988
A:Title: Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.
A:Reference number: J0345; MUID:89137980; PMID:2976010
A:Accession: J0345
A:Molecule type: DNA
A:Residues: 1-1375 <USD>
A:Cross-references: UNIPROT:P13470; UNIPARC:UPI0000155515
A:Experimental source: GS-5
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
J. Bacteriol. 169, 4263-4270, 1987
A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.
A:Reference number: A33135; MUID:87308013; PMID:3040685
A:Accession: C33135
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-349 <SHI>
A:Cross-references: UNIPARC:UPI000014E25D; GB:M17361
C:Genetics:
C:Function:
A:Description: catalyzes the synthesis of both water-soluble and water-insoluble glucan:
C:Keywords: duplication; glycosyltransferase; hexosyltransferase
F:1-14/Domain: signal sequence #status predicted <SIG>
F:35-1375/Product: glucosyltransferase #status predicted <MAT>
F:1126-1145/Domain: cpl repeat homology <CP1>
F:1253-1272/Domain: cpl repeat homology <CP2>
F:1318-1337/Domain: cpl repeat homology <CP3>

Query Match 100.0%; Score 7143; DB 2; Length 1375;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MEKKVRFKLKVKRWTVTSIASAVVTLTSLSGSLVKADSTDDROQAVTESQASLVTTSE	60
DB	1	MEKKVRFKLKVKRWTVTSIASAVVTLTSLSGSLVKADSTDDROQAVTESQASLVTTSE	60
QY	61	AAKETLTATDTSTATSQPTATVTDNVSTTNQSTNTTANTANFVVKPTTTTSEQAKTDN	120
DB	61	AAKETLTATDTSTATSQPTATVTDNVSTTNQSTNTTANTANFVVKPTTTTSEQAKTDN	120
QY	121	SDKIITTSKAVNRLTATGKFPANNNTAHPKTVTDKIVIPKPKIGKLPQSSLSQDDIAA	180
DB	121	SDKIITTSKAVNRLTATGKFPANNNTAHPKTVTDKIVIPKPKIGKLPQSSLSQDDIAA	180


```
QY 831 RVAASTAPSTDCSKSVHONALDSRVMPFEGSFNFQAFATKBEYTNVIAKNVDKEAEKGV 890
Db 802 RVAASDTASTDCSKSLHQDAAMDSRVMPFEGSFNFQAFATKBEYTNVIAKNVDKEAEKGV 861
QY 891 TDFEMAPQVYSTDCSGFLDSVQNGYAFTRDYLIGISKPKNKYGTADDLVKAIKALHSHSGI 950
Db 862 TDFEMAPQVYSTDCSGFLDSVQNGYAFTRDYLIGISKPKNKYGTADDLVKAIKALHSHSGI 921
QY 951 KVMADWVDPQMYALPEKEVVTATRVVDKYGTPVAGSQIKNTLVYVDGKSGKDDQAKYGA 1010
Db 922 KVMADWVDPQMYALPEKEVVTATRVVDKYGTPVAGSQIKNTLVYVDGKSGKDDQAKYGA 981
QY 1011 FLEELQAKYPELFAKQISTGVPMPSVKIKQWSAKYFNGTILGRGAGYVLKDAQNTY 1070
Db 982 FLEELQAKYPELFAKQISTGVPMPSVKIKQWSAKYFNGTILGRGAGYVLKDAQNTY 1041
QY 1071 FSLVSDNTFLPKSLVNPNGHTSSVTLVFDGKGVVYVYST-SGNQAKNAFISLGNWYVF 1129
Db 1042 FNASDITFLPSSLLG-----KWESGRYDGGKGIYNSATGDDQVKASFITEAGNLVYF 1096
QY 1130 DNNGYMTGAOSINGANYFSLNGIQLRNAIYDNGKNVLSYYGNDGRRY--ENGYLFGQ 1187
Db 1097 GKGGMVTVGAQTINGANYFSLNGIQLRNAIYDNGKNVLSYYGNDGRRY--ENGYLFGQ 1156
QY 1188 QWRYPQNGIMAVGLTRVGAQVYFASGFAKQGFITTAGDKLRYFDRSDSGNOISNRVFR 1247
Db 1157 DWRYPQNGIMAVGLTRVGAQVYFASGFAKQGFITTAGDKLRYFDRSDSGNOISNRVFR 1216
QY 1248 NSKGWELFDHNGVAVTGTVTFNGQRLYFKPNQVQAKGEFTDANGYLRYDPSNGNEVR 1307
Db 1217 DKTGHVYLGKDGVAVTGAQTVGKQKLYFEANGQQVKGDFVTSDEGKLYFYFDVDSGDWMT 1276
QY 1308 NRVFNSKGEWFLFDHNGIATVGARVANG 1336
Db 1277 DTFIEDKAGNWFYLGKDGAAVTGAQTIRG 1305

RESULT 4
dextranase (EC 2.4.1.5) precursor - Streptococcus mutans
C:Species: Streptococcus mutans
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Mar-2004
C:Accession: A45866
J.Honda, O.; Kato, C.; Kuramitsu, H.K.
J. Gen. Microbiol. 136, 2099-2105, 1990
A:Title: Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the glucosyl
A:Reference number: A45866; MUID:91100958; PMID:2148600
A:Accession: A45866
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-1431 <HON>
A:Cross-references: UNIPARC:UPI000017AC5C; GB:M29296
C:Keywords: glycosyltransferase; hexosyltransferase
F:181-201/Domain: cpl repeat homology <CP1>
F:1127-1146/Domain: cpl repeat homology <CP2>
F:11192-1211/Domain: cpl repeat homology <CP3>
F:1257-1276/Domain: cpl repeat homology <CP4>
F:1277-1297/Domain: cpl repeat homology <CP5>
F:1321-1340/Domain: cpl repeat homology <CP8>
F:1341-1361/Domain: cpl repeat homology <CP6>
F:1385-1404/Domain: cpl repeat homology <CP7>

Query Match 50.3%; Score 3591.5; DB 2; Length 1431;
Best Local Similarity 51.9%; Pred. No. 4e-169;
Matches 721; Conservative 215; Mismatches 350; Indels 103; Gaps 22;

QY 1 MEKKVRFLKRVKRWVTVSIAAVVTL-TSLGSLVKAADS-----TDRQQA 47
Db 1 MEKKRYKMKHVKRWVTVSIAAVVTL-TSLGSLVKAADS-----TDRQQA 60
QY 48 VTESQASLVTTSEA-AKETLTATDTSTATSQPTATVTDNVSTNOS-----TNTTAN 101
Db 61 SESSQTPAKTKQAQTEQTAQSQANVADTSITKEFSPQNIITQANSDDKTVNTKSE 120
```

```
QY 102 TANFVVKPTTTSQAKTDNSDKIITTSKAVNRLTATGKFPANNNTAHPKTVTDKIVPIK 161
Db 121 EAQTSERTKQSEAOITASSQALTOAKA--ELT----- 152
QY 162 PKIGKLGKOPSSLSQD-----DIAALGNVKNIRKVNKYYYKEDGTLQKNVALNINGKTF 216
Db 153 -----KORQTAAGENKNPVDLAI PNK---QIDGKYYYIGSDQOPKPNFALTNNKVL 203
QY 217 FFDE-TGALSNNITLPSKKGNIT--NNDNTNSFAQVQVYSTDVANFEHVDHYLTAESWYR 273
Db 204 YFDKNTGALTDTSOVQFKGLTKLND---YTPHNQIVNFEPTSLETIDNVTADSWYR 259
QY 274 PKYILKQKWTQSTQKDFRPLMTWPDQETQKQVYVYMAQ-LGIHQTYNTATSPLOL 332
Db 260 PKDILKNGKWTASSESDLRPLMSWPDQKQIAYLNNMQQGLGTGENYATSSQESL 319
QY 333 NLAQTIQTKTEKITAFKNTNWLROTISAFVKTSQAMNSDSEKPPD---DHLQKCALL 388
Db 320 NLAQTVQVKLTETKISQTOQLRDIINSFVKTPWNNSQTESDTSAGEKDHLOGGALL 379
QY 389 YSNNSKLTQSANRYRILNRTPTNQTKGKDPRTYADRTIGGYEFLLANDVNSNPVQAE 448
Db 380 YSNSDK-TAYANSYRLLNRTPTSTQTK--PKYPEDNSSGGVDFLLANDIDNSNPVQAE 436
QY 449 QLNMLHFLMNFNGNIYANDPDANFDSIRVDADVNDVADLLQIAGDYLLKAAKGHKKDKAAN 508
Db 437 QLNMLHLYMNGSIVANDPEANFDGVRVDADVNDVNADLLQIASDYLLKAHYGVYKSEKNAI 496
QY 509 DHLSTLEAWSYNDPYPYLDHDDGNMNMNRLRLSLYSIAKPLNO-----RSGMNP 560
Db 497 NHLSTLEAWSNDPOYNKDKTGAQPIDNKURLSLLYALTRPLEKDAENKNEIRSGLEPV 556
QY 561 ITNSLVNRDNDNAETAAPVPSFIRAHDSYVDQLIRNIIRTEINPNVVGYSFTTEIKKA 620
Db 557 ITNSLNNRSEAGKNSERMANYIFIRAHDSYVQTVIAKIIKAQINPKTDGLTFLDELQQA 616
QY 621 FEIYNKOLLATEKKYTHYNTALSYALLTNKSSVPRVYVYGDFTDDGQYMAHKTINYEAI 680
Db 617 FKINEDMRQAKKYTQSNIPTAYALMLSNKDSITRLYGYDMYSDDGQYMATKSPYDAI 676
QY 681 ETLKARLKYVSGQAMENQOVGNSE-----IITSVRYGKALKATDGTDRTRTS 731
Db 677 DTLKARLKYAAGQDMKITTVGDKSHMDWDYTGVLTSVRYGTGANRATDQSSATKIQ 736
QY 732 GVAVTEGNPNRLKASDRVVVNNGAAHKNQAYRPLLLTTDNGIKAYHSDQEAAGLVRYT 791
Db 737 GMAVITSNNPSLKLNDQKVIIVNMGAAHKNQAYRPLLLTTKDGLTSTYSDAAAKSLYRKT 796
QY 792 NDRGELIFTAADIKGYANPOVSGYLGWVVPVGAADQDVRVNAASTAPSTDGKSVHONAL 851
Db 797 NDKGELVFDASDIQGYLNPQVSGYLAVVWVPGASDNQDVRVAASNKANATGOVYSSSAL 856
QY 852 DSRVMEGFSNFOAFATKKEEYTNVIAKNVDKPAEWGVTDFEMAPQVYSSVSDGFLDSV 911
Db 857 DSQLIYEGFSNFQDFVTIKDSITNKKIAQNQLFKSWGVTSEFEMAPQVYSSDGSFLDSI 916
QY 912 IQNGYAFTRDYLIGISKPKNKYGTADDLVKAIKALHSHSGIKVMADWVDPQMYALPEKEVVT 971
Db 917 IQNGYAFEDRYDLAMSKNKYGSQQDMINAVKALHSHSGIQVIADWVDPQIYNLPKGVVT 976
QY 972 ATRVDKYGTPVAGSQIKNTLVYVDGKSGKDDQAKYGAFLPELQAKYPELFAKQISTG 1031
Db 977 ATRVNDYGEYRKDSIKNTLYAANTKSNKGQYQAKYGGAFLSLAAPYSIFNRQTQISNG 1036
QY 1032 VPMDSVKIKQWSAKYFNGTILGRGAGYVLKDAQNTYFSLVSDNTFLPKSLVNPNGHT 1091
Db 1037 KKIDSEKITAWKAYFNGTILGRGAGYVLKDAQNTYFSLVSDNTFLPKSLVNPNGHT 1091
QY 1092 SSSVTGLVFDGKGVYVYSTSGNQAKNAFI-SLGNWYVFDNNGYVMTGAQISNGANYYP 1150
Db 1092 KEASTGFVNDGNGMTFYSTSGVQAKNSFVQDAKGNWYVFDNNGHMYVGLQQLNGEVOYEL 1151
```

474	LNNS-ATSHANSDFRLMNRFTTQGTGR--KYHIDRNSGGYELLANDINDSNPNPAQAEQ	530
450	LNWLHFLMNFNIYANDPANFDSIRVDADVNDVADLLQIAGDYLKAAGHKHNDKAAND	509
531	LNWLHYIMNIGSILGNDFPSANFQGVRIADVNDVADLLQIASDYFKEKYRVADNEANAIA	590
510	HLSTLEAWSYNDTPYLHDDGNMIMNDRLRLSLYLAKPLNQORSQMNPLTINSLNVRT	569
591	HLSTLEAWSYNDHQYNKDTKGAQLSDINPLRETIILLTFLRKSNYGRSLERVITNSLNNRS	650
570	DDNAETAAVPSYSPIRAHSDSEVDLIRNIIRTEINPNVUGSYSTTEBEIKKAFETYNKDLL	629
651	SEQKHTPRDANIIPVRAHDSQVAVLANIISKQINPKTDGTFITMDSLKQAFETYNADIA	710
630	ATEKKYTHYNTALSYALLTNKTSVPRVYDGMFTDDGQYMAHKNTINYEAIETLLKARIK	689
711	KADKKYTOYNIIPAAYATMLTNKOSITRVYGDLFITDDGQYMAEKSPPYNAIDALLRARIK	770
690	YVSGGQAWRNOQVGNSEIITSVRYGKGALKATDTGDRTRTSTGVAVIEGNNPSLRLKASD	749
771	YVAGGQMKVTKLNGYEIMSVRYGKGAEBANQLGTAETRNQGMVLVTANRPDMWKLGCAND	830
750	RVVYNNGAHKNQAYRPLLLITTINGIKAYHSQDEA-AGLRYVYNDRGCELLFTAADIKGYA	808
831	RLVYNNGAHKNQAYRPLLLSKSTGLATYLKQSDVPAGLVRYTDDQGNLFTTADDIAGHS	890
809	NPOVSGYLVGVWVPVGAADODVVAASTAPSTDGKSVHQ--NAALDSRVMEFGSPNFQAEA	867
891	TVEVSGYLAVWVPVGAASENQDARTKASS--TKKGQVFPSSAALDSQVITYEGFSNPDFV	948
868	TKKEGYTNVVIKXNDKFAEWGVTDEMAPQYSSSTDGSLDSVIONGYAFTDRYDLGIS	927
949	KTPSQYNTVIAQNAKLFKEWGITSEFEAQQYVSSQDGTFLDSIENGYAFEDRYDIAMS	1008
928	KPNKYGTADDLVKAIKALHSKGIKWADWVPDQMYALPEKEVVTATRVDKYGTFPVAGSQI	987
1009	KNNKYGSLKQMLDALRALHAEGISAIADWVPDQIYNLPKGVEVVTASRTNSYGTPRNABE	1068
988	KNTLYVVDGKSGKQOAKYGGAFLELOAKYPFLPARKQISITGVMPDPSVKIKQMSAKY	1047
1069	YNSLYAAKTRTFGNDPGKYGGAFLELKAAPAIERVOISNGRKUUTTNKEKITQMSAKY	1128
1048	FNGNTILRGAGYVLKQDQATNTYFSLVSDNTFPLKSLVNPNHGTSSTSVTLGVFD--GKGY	1105
1129	FNGSNIQGTGARVVLQDNATNQYFSVKAGQTELPKQM-----TEITGSGFRRVGGDDV	1180
1106	VYTSYTSNQAKNAFISLIG-NNWYTFDNNQYFWVTGAQSGINGANYTFYLSNGIQLRNAYDNG	1164
1181	QYLSIGYGLAKNTFIOVGANQWYTFYFKNGNMNVTEQVIDGKKYFFFDLNGLQLRLHRQGS	1240
1165	NKVLSYDNGDRYENGYYLFG--QOWRYFO-NGIMAVGLTRVHGAVQYFD-ASCGFOAK	1219
1241	DGHVYYDYPKGVAQFNFGYDFAGPRQDVRFDGNGQYRGLHDMYGTTFYFDEKTIQAK	1300
1220	GQFITTTADGKRLRYFDRSDSGNQISNRFRVNRKGS-WFLFDHNGVAVTGTVTFNGQRLYFKP	1278
1301	DKFIRFADGTRFYIPDTGNLAVNRPAQNENKAWYLLDSNGYAVTGLQTINGKQYYFDN	1360
1279	NGVOAKGEFTRDANGYRLYY-DPNSGNEVNRFRVNRKGSWFELFDHNGIAVTGARVVNGH	1337
1361	EGROVKGHFYTIINN--ORVFLDGDSGEIASPRFV-TENNKWYVVDGNGKLKVGQAOVINGN	1417

RESULT 6
T30857
glucosyltransferase - Streptococcus salivarius
C:Species: Streptococcus salivarius
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30857
F:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.
Infect. Immun. 63, 609-621, 1995
A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes c
A:Reference number: Z20909: MUID:95122197: PMID:7822030

Query Match	45.0%;	Score 3216;	DB 2;	Length 1577;
Best Local Similarity	47.7%;	Pred. No. 1.6e-150;		
Matches 687;	Conservative 214;	Mismatches 413;	Indels 126;	Gaps 33
Qy	1	MEKVRFKLRKVKKQWTVTSASAVVLTLSLG-----SLVKA DST-----DDRQAVT	49	
Db	1	MENKVRFKLHKVKKNVTVIGTV--LSMVLAGGSLLAQGVKVEADETSAPNGDGLQQLSE	58	
Qy	50	ESQASLVTTTS-----EAAKETLTATDTST-----ATSATSO-----PTAT--	84	
Db	59	DGTASLVTTTTVTEQASQAQSAVSAVATASVSHETSFQAATSASQBATQAQTSVPASOE	118	
Qy	85	--VTDNVSTTNQSTNTT-----ANTANFVVKPT-----TTSQAK-----TDNSDKII	128	
Db	119	VAVSSQTQSSQGETQTTEVSSQGTSTVAGQTSQAQSTPSVTEQARPRVLITNAAIAIATR	178	
Qy	129	KA-----VNRLTATGKFPVANNHTAHPKTVTDKIVIPKPI-----	164	
Db	179	AADSTIRINANRNTNITITASGTTENVNITIGPNTP-KPNVTVTPSGNTRPNVTIVTQPN	237	
Qy	165	--GKLQBPSSLSODDIAALGN-----VKNIRKNGKYYIYKEDGTLOKRYALNINNG	213	
Db	238	QPNKPVQSPQSPQPNKVPQNPQPSLDYKFPVASNLKTIIDGKQYIV-ENGWVKKNAI	296	
Qy	214	KTFEDEFETGALSNNITLPSKGN-ITNNDNTNSFAQYNQVYSTDVANFEHVDHYLTAESWY	272	
Db	297	RLYYFDETGAVMDQSKPYRADAI PNN---SIYAVYNQAYDTSSKSFHDLNDFLTADSWY	353	
Qy	273	RPKYLKDGKWTQSTEKDFRPLLTWMPDQSTQRYVNYMNAQLGIHOTYNTATSPQLQ	332	
Db	354	RPKQLKDGKRWTASTEKDYRPLLTWMPDKVTQVNYLNTYMSQQQFGNKTYTDDMMSYDL	413	
Qy	333	NLAAQTIOTKIEEKITAENKTNLWRLQTISAFVKTQSANWSNSEKPF--DDHLQKCALLY	389	
Db	414	AAAETVQGIIEERIGREGNTTLWRLQMSDFIKTQPGWNSESDNLLVCKHLOQCALTF	473	
Qy	390	SNNKSLTQANSYRILNRPTPNTQTKGKDPRYATDRTIGGYEFFLLANDVDNSNPVVAEQ	449	

A:Accession: T30857
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1449 <SIM>
A:Cross-references: UNIPROT:Q55264; UNIPARC:UPI00000B166E; EMBL:L35495; NID:G662378; PID
C:Genetics:
A:Gene: gtfL

Query Match 44.2%; Score 3159; DB 2; Length 1449;
Best Local Similarity 47.0%; Pred. No. 8.8e-148;
Matches 669; Conservative 197; Mismatches 439; Indels 118; Gaps 28;

QY 1 MEKKYRFLKRYKRWTVSI-----ASAVVTLTSLGSLVKADSTDDROOAVTESQAS 54
DB 1 MDKKVHYKMKVKQWVIAVGLSGLGAVSAVSLGTNDG-VVQADEHTDATVAIPD---I 56
QY 55 LVTTSEAAKETLTATDTSTATSQ---PTATVTDNVSTTNQSTNTTANTFVVKPTT 111
DB 57 TVDTGTGVSNDTTAAQDPTTAAVATNDVATQATPTATPDLTDTTNTVAANAADVTATVG 116
QY 112 TSEQAKTNSDKIITTSKAVNRLTATGKFPVANNNTAHPKTVTD----- 155
DB 117 TDRAATT--NDTATNDTAVDTTN-----NNTTDTTVDRAATTERRATGARRGP 166
QY 156 ----KIVPIKPKIGKLPSSLSQDDIAALGNV-----KNIRKVGKYYKKEDGTLOKN 206
DB 167 TGGRRATPVNGTNNANNNTVTVVNDLPATNNVTDGPGSHIKTINGKQYVVEDDGTIRKN 226
QY 207 YAL-NINGKTPFFDETALSN-----NTLPSKGNITNNDNTNSFAOYNQVY 252
DB 227 YVLERIGGSQYFNAETGELSNGKEYRFDKNGGTGSSADSTNTNTVTVGDKNAF-----Y 280
QY 253 STDVANFEHVHYLFAESWYRKYILKOGKWTQSTQTEKDFRPLMTWPPDQSTQYVNY 312
DB 281 GTTKDIELVDGYFTANTWYRKEILKDGKEWTASTENDKRPLLTVMWPSKAIQASLYN 340
QY 313 MNAO-LGHTQNTATSPQLNLAQTIQTKLEKITAENKTNMLRQTISAFVKTSQAWN 371
DB 341 MKEQGLTNGTQTSFSQTMQDALEVKRLERAREGNTDMLRTIKNFVKTPQGN 400
QY 372 SDSEK-PFDDHLQKGLLYSNNSKLTQSANSYRILNRTPTNTQTKGKDPRYTADRTIGY 430
DB 401 STSENLDNNDHLQGGALLYNDSR-TSHANSYRLLNRTPTSTQTKHNPKYTKDTSNGGF 459
QY 431 EFLANDVNSNPVQAEQLNLHFLMNFNTYANDPDANFDSIRVDVNDVADLLQIA 490
DB 460 EFLANDIDNSNPAVQAEQLNLHLYIMNIGTTTGGSEDEFDGVVRVDAVDNVDNADLLQIA 519
QY 491 GDLAKAAGIHKNDKAANDHLSILEAWSYNDPTPYLHDDGDNMNDNRLRLSLLYSLAKP 550
DB 520 SDYFKAKYGADQSQQAIKHLSILEAWSHNDAYNEDTKGAQLPMDDDPHLALVYSLLRP 579
QY 551 LNQRSGMPLITNSLVNRTDDNAETAAPVSYSFIRAHIDSEVQDLIRNIIRTEINPNVVG 610
DB 580 IGRSGVPELISLNSLDRSECKNSKRWANAYAFVRAHIDSEVQSIIQIINKNEINQSTGN 639
QY 611 SPTTEIKAFIYINKDLATKKYTHYNALSYALLTNKSSVRVYVYGMFTDDGQVM 670
DB 640 TPTLDEMKAFAIYNKDMESANKQYTVNIPSAVALMLTHKDTVPVRYVYGMFTDDGQVM 699
QY 671 AHKTINYEAIETLLKARIKYSGGOAMRNOQVGNSE-----IITSRYCKGALKATD 722
DB 700 AQKSPYDAIETLLKRIYRAGGQDMKVNYIGYNTNGWDAGVLTSTVRYGTGANSASD 759
QY 723 TGDRTTRTSGVAVIBGNPSPRLKASDRVVNNGAAHKNQAYRPLLLTTDNGIKAYHSDQ 782
DB 760 TGAETRNQGMVIVSNQPALRL--TNSLTINWGAHRNQAYRPLLLTTDNGVATYLNDS 817
QY 783 EAAGLVRYTNDREGELIFTAADIKGVANPOVSGYLGWVPVGAADQDVRVAASTAPSTDG 842
DB 818 DANGIVKYTGDGNGLTFTSANEIRGNPQVDGYLAVWVPVGAENQDVRVAPSEKNSGS 877
QY 843 KSVHQAALDSRVMEFGFSNFQAFATKKEEYTNVVIKAVNDKFAEWGVTDPMAQYVSS 902

DB 878 LVYESNAALDSQVIYEGESNFQDFVQNPSTQNTKKIAENANLFSKSGITSEFAQYVSS 937
QY 903 TDGSLDSVIONGYAFTDRYDLGISKPNKYGTADDLVKAIKALHSKIGIKVMADWVPDQMY 962
DB 938 DDGSLDSVIONGYAFTDRYDIGSKDNKYGSLADLKAALKSLHVAIGISAIADWVPDQYI 997
QY 963 ALPEKEVVTATRVDKYGTVPVAGSQIKNTLYVVDVSSGKQDOAKYGGAFLEBLQAKYPEL 1022
DB 998 NLPGEVVVATRVNNYGETKDGAIIDHSLYAAKTRTFGNDYQGGYGGAPFLDELKRLYPOI 1057
QY 1023 FARKQISTGVPMWPSVKIKQWSAKYFNGTNIILGRGAGYVLKDQATNTYFSLVSDNTFLPK 1082
DB 1058 FDRVQISTGKMTTDEKITQWSAKYMGNTNILDGSEVVLKN-GLNGYGTNGGKVSUPK 1116
QY 1083 SLVNPNHGTS-SSVTGLVFDGKGYV-----YYSTSGNOAKNAFISLGN-NWYFDNN 1132
DB 1117 -VGSNQSTNGDNQNG--DGSCKPEKRLFSVRVRYNNGQYAKNAFIKDNDGNVYFDS 1172
QY 1133 GYMTGAOSINGANYFSLNGIOLRNALYDNGNKVLSVYGNDRRYENG----- 1181
DB 1173 GRMAVEKTIIDGQYFFLANGVQLRDGYRQRRRGQVFFYDQNGVLNANGKQDPKPDNNNN 1232
QY 1182 -----YYLFGQQ-WRYEQ-NGIMAVGLTRVHGAVQYFDASGFOAKGOFITTAGDKLRYF 1233
DB 1233 ASGRNQFVQIGNNVWAYDDGNGKRVTHQNINGQELPFDNNGVQVKGRTV-NENGAIYY 1291
QY 1234 DRDSGNQISNFRVNSKGEWFLFDHNGVAVTGTVTTFNGORLYFKPNGVQAKGEFIRDANG 1293
DB 1292 DANGEMARNRFAETEPGVMAYFNNDGTAVKGSQINGQDLYFDQNGRQVKA-LANVDG 1350
QY 1294 YIRYVDNPSGNEVRNRFVNSKGEWFLFDHNGIATVGARVNG 1336
DB 1351 NLRYVDNBSGELYRNRF-HEIDGSIYFDGNGNAVKGVMWNING 1392

RESULT 7
T30552
glucosyltransferase N - Streptococcus salivarius (fragment)
C:Species: Streptococcus salivarius
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30552
R:Jaffe, R.I.
submitted to the EMBL Data Library, February 1998
A:Description: Streptococcus salivarius VI477 gtfN.
A:Reference number: Z20854
A:Accession: T30552
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1449 <JAF>
A:Cross-references: UNIPROT:Q68542; UNIPARC:UPI00000B10FD; EMBL:AF049609; NID:G29335545;
C:Genetics:
A:Gene: gtfN

Query Match 44.0%; Score 3146; DB 2; Length 1449;
Best Local Similarity 46.8%; Pred. No. 3.9e-147;
Matches 666; Conservative 200; Mismatches 439; Indels 118; Gaps 28;

QY 1 MEKKYRFLKRYKRWTVSI-----ASAVVTLTSLGSLVKADSTDDROOAVTESQAS 54
DB 1 MDKKVHYKMKVKQWVIAVGLSGLGAVSAVSLGTNDG-VVQADEHTDATVAIPD---I 56
QY 55 LVTTSEAAKETLTATDTSTATSQ---PTATVTDNVSTTNQSTNTTANTFVVKPTT 111
DB 57 TVDTGTGVSNDTTAAQDPTTAAVATNDVATQATPTATPDLTDTTNTVAANAADVTATVG 116
QY 112 TSEQAKTNSDKIITTSKAVNRLTATGKFPVANNNTAHPKTVTD----- 155
DB 117 TDRAATT--NDTATNDTAVDTTN-----NNTTDTTVDRAATTERRATGARRGP 166
QY 156 ----KIVPIKPKIGKLPSSLSQDDIAALGNV-----KNIRKVGKYYKKEDGTLOKN 206
DB 167 TGGRRATPVNGTNNANNNTVTVVNDLPATNNVTDGPGSHIKTINGKQYVVEDDGTIRKN 226

703 FEAPQYRASDKSFLDAIVQNGYAFTRDYDIGYNTPTKYGTADNLLDALRALHGGGIA 763

953 MADWVPDMQYALPEKEVVTATRVDKYGTVPVAGSOIKNTLYVVDGKSGKQDQAKYGGAF 1011

763 INDWVPDQIYNLPDEQLVTAIRTDGSDHTYGSVIDHTLYASKTVAGGIYQO-QIYGAF 821

1013 EELQAKYPELFAKQISTGVPMDPVSKIKOWSAKYFNGTNIILGRGAGYVLKDQATNTYFS 1077

822 EQLKTYQPLFOQKOISTDQPMNPDIIQIKSWEAKYFNGSNIQGRGAWYVLKDWGTOQYFN 881

1073 LVSNTETPLKSLVNPNGHGTSSVTCGLVFDGKGYVYYSSTGSGNOAKNAFISLGNWYTFDNN 1133

882 VSDAQTEFLPKOLLG-----EKAKTGFVTRGKETSFYSTSGYQAKSAFICDNGNMYFFDDK 936

1133 GYMTGAQSGINGANYEFLSNGIQLRNA--IYDNGMKVLSYVYNGDGRYVYENGYYLFGQO-- 1188

937 GKAVVNGQVINGINYELFNGIELQDAYLVHDG---MYYVYNTIGKQLHNTYYQDKQKNF 993

1189 WRYPQNGIMAVGLTRV-----HGAVQYFDASGFOAKGQIFITADGKLYRYFD----- 1233

994 HYFFEDHMAQGIWIIQSDGTPVTQYFDENGKQKGVAVKSGDGLHYFDGASGNMLF 1051

1235 -----RDSGNQISNRFRVNSKGWFLFDHNGV 1266

1054 SWGLRADGSLWLYDBEKGNAVTKQTIINNQTVYFNDGGRQIKKNFKELADGSLWLYLNKGV 1111

1262 AVTGTVTNFGRLYFKFNGVQAKGEFTRDANGYLRYDPPNSGNEVRNRFVNSKGWFLF 1321

1114 AVTGEIINGQTLFNGDGRQFKGTTTHNATGESRYDDPSDGNMITDRFERVDGNQWAYF 1177

1322 DHNGIAVTGARVVNG 1336

1174 GYDGVAVTGRIIKG 1188

RESULT 13

B97033

Characterized protein, related to enterotoxins of other Clostridiales [import

C;Species: Clostridium acetobutylicum

C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004

C;Accession: B97033

R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum ATCC824

A;Reference number: A96900; MUID:21359325; PMID:21359325

A;Accession: B97033

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-2817 <KUR>

A;Cross-references: UNIPROT:Q97K42; UNIPARC:UP100000CA0A0; GB:AE001437; PIDN:AM

A;Experimental source: Clostridium acetobutylicum ATCC824

C;Genetics:

A;Gene: CAC1079

Query Match 6.0%; Score 428.5; DB 2; Length 2817;

Best Local Similarity 21.0%; Pred. No. 1.2e-12;

Matches 359; Conservative 182; Mismatches 530; Indels 637; Gaps 94

34 SLVKADSTDRQAVTESQASLVTTSEAKETLTATDTSTATSATQPTATVTD-NVSTT 92

118 SIPVTNVTDNKTNFKNESSIN---NEA---PIIPKDTSKTSSTSAQTGSDNNNIPSN 170

93 NQSTNTT--ANTANFVVKPTTTSEQA-----KTDNSDKIITTSKAV--NRUTA-----TGK 139

171 NTSNTSNTKSNPNSNTDIKTTEAPANAPIKOTPNQSDSALAKXKALGNLNLADSSQTSK 230

140 FVPANNNTAHPK--TVTDK-----IVPIKPK----- 163

231 VTSSNDA--PKVNTTSTDKAANLNDSDQGWVTGDKKYYVNGVQKGFQSGINKSIY 288

164 -----IGKLKQPSLSQDD-----IAALGNVK-----NIRKVNKYY 195

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.
OM protein - protein search, using sw model
Run on: February 11, 2006, 18:57:35 ; Search time 185.535 Seconds
(without alignments)
5228.676 Million cell updates/sec
Title: US-10-797-821-35
Perfect score: 7143
Sequence: 1 MEKKVRFKLKRVKRWTVS.....VKVNTWILIPMKFVIM 1375

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 2166443 seqs, 705528306 residues
Total number of hits satisfying chosen parameters: 2166443
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : UniProt_05.80.*
1: uniprot_spot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	6905	96.7	1455	1 GTFB_STRMU
2	5350	74.9	1476	1 GTFB_STRMU
3	4214.5	59.0	1590	2 Q59983_9STRE
4	4207.5	58.9	1590	2 Q55263_9STRE
5	4194	58.7	1597	1 GTF1_STRDO
6	4168.5	58.4	1592	1 GTF2_STRDO
7	3579.5	50.1	1462	1 GTFD_STRMU
8	3366.5	47.1	1577	2 Q54178_STRGN
9	3348.5	46.9	1454	2 Q69A94_LEUME
10	3293.5	46.1	1575	2 Q9LCH3_STROR
11	3216	45.0	1577	2 Q55265_STRSL
12	3197.5	44.8	1512	2 Q9WXJ5_9STRE
13	3175.5	44.5	1506	2 Q56CX8_9STRE
14	3161.5	44.3	1554	2 Q8KZL5_9STRE
15	3159	44.2	1449	2 Q55264_STRSL
16	3146	44.0	1449	2 Q68542_STRSL
17	3111	43.6	1561	2 Q58BM0_9LACO
18	3081.5	43.1	1527	2 Q8KRE1_LEUME
19	3077.5	43.1	2835	2 Q8G9Q2_LEUME
20	3070	43.0	1518	2 Q00600_STRSL
21	3069.5	43.0	1522	2 Q5TXV4_LEUME
22	3065.5	42.9	1527	2 Q9ZAR4_LEUME
23	3044	42.6	1508	2 Q52224_LEUME
24	3031	42.4	1508	2 Q9EZHS_LEUME
25	3026.5	42.4	1477	2 Q9L466_LEUME
26	2970	41.6	1330	2 Q84CN4_LEUME
27	2899	40.6	1599	2 Q00599_STRSL
28	2893.5	40.5	1365	1 GTF5_STRDO
29	2876.5	40.3	1290	2 Q48756_LEUME
30	2861	40.1	1595	2 Q58BM3_LACSK
31	2836	39.7	1338	2 Q9WXJ4_9STRE

32	2770	38.8	1463	2 Q58BM6_LACFE
33	2501.5	35.0	2057	2 Q9RE05_LEUME
34	2487	34.8	591	2 Q8VUH3_STRMU
35	2481	34.7	1016	2 Q9LCJ7_LEUME
36	2403	33.6	1772	2 Q58BN0_LACRE
37	2383.5	33.4	1772	2 Q58BN3_LACRE
38	2356.5	33.0	1781	2 Q4UCS4_LACRE
39	2351.5	32.9	1781	2 Q58BL9_LACRE
40	2351.5	32.9	1781	2 Q4JLC7_LACRE
41	1542	21.6	1619	2 Q58BM0_LACRE
42	1529	21.4	1231	2 Q58BN1_LACRE
43	1023	14.3	522	2 Q8VV10_STRSA
44	848.5	11.9	374	2 Q6ZX19_9LACO
45	428.5	6.0	2817	2 Q97K42_CLOAB

ALIGNMENTS

RESULT 1
GTFB_STRMU STANDARD: PRT: 1455 AA
AC P13470; O69382; O69385; O69388; O69391; O69397; P05427;
DT 01-NOV-1988 (Rel. 09, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Glucosyltransferase-SI precursor (EC 2.4.1.5) (GTF-SI).
DE (Dextranucrase) (Sucrose 6-glucosyltransferase).
GN Name=gTFC; OrderedLocusNames=SMU.1005;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]_TaxID=1309;
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=GS-5;
RX MEDLINE=89137980; PubMed=2976010; DOI=10.1016/0378-1119(88)90382-4;
RA Ueda S., Shiroza T., Kuramitsu H.K.;
RT "Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.";
RL Gene 69:101-109(1988).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MT4239 / Serotype C, MT4245 / Serotype e, MT4251 / Serotype f, MT4467 / Serotype e, and MT8148 / Serotype c;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S., Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
RN [3]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=UAI59 / ATCC 700610 / Serotype c;
RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;
RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J., Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P., Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A., Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
RN [4]
RP NUCLEOTIDE SEQUENCE OF 1-349.
RC STRAIN=GS-5;
RX MEDLINE=87308013; PubMed=3040685;
RA Shiroza T., Ueda S., Kuramitsu H.K.;
RT "Sequence analysis of the gtfB gene from Streptococcus mutans.";
RL J. Bacteriol. 169:4263-4270(1987).
CC -!- FUNCTION: Production of extracellular glucans, that are thought to play a key role in the development of the dental plaque because of their ability to adhere to smooth surfaces and mediate the aggregation of bacterial cells and food debris.
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl)(n) = D-

```

CC CC fructose + (1,6-alpha-D-glucosyl)(n+1).
CC CC -!- SUBCELLULAR LOCATION: Secreted.
CC CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
CC CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes
CC CC water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
CC CC forms of glucans.
CC CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
CC CC -!- SIMILARITY: Contains 5 cell wall binding repeats.
CC CC -----
CC CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use as long as its content is in no way modified and this statement is not
CC CC removed.
CC CC -----
CC CC EMBL; M22054; AAA88592.1; -; Genomic DNA.
CC CC EMBL; D88652; BAA26102.1; -; Genomic DNA.
CC CC EMBL; D88655; BAA26106.1; -; Genomic DNA.
CC CC EMBL; D88658; BAA26110.1; -; Genomic DNA.
CC CC EMBL; D88661; BAA26114.1; -; Genomic DNA.
CC CC EMBL; D89978; BAA26120.1; -; Genomic DNA.
CC CC EMBL; AE014940; AAN58706.1; -; Genomic DNA.
CC CC EMBL; M17361; AAA88589.1; -; Genomic DNA.
CC CC PIR; J70345; J70345.
CC CC HSP; P06653; IH8G.
CC CC InterPro; IPR002479; Cell_wall_bd_put.
CC CC InterPro; IPR003318; Glyco_hydro_70.
CC CC Pfam; PF01473; CW_binding_1; 2.
CC CC Pfam; PF02324; Glyco_hydro_70; 1.
CC CC Complete proteome; Dental Caries; Glycosyltransferase; Repeat; Signal;
CC CC Transferrase.
CC CC -----
CC CC SIGNAL 1 34
CC CC CHAIN 35 1455
CC CC REPEAT 1126 1159 Glucosyltransferase-SI.
CC CC REPEAT 1126 1200 A repeat.
CC CC REPEAT 1227 1238 A repeat.
CC CC REPEAT 1253 1303 C repeat.
CC CC REPEAT 1318 1330 AC repeat.
CC CC REPEAT 1318 1330 A repeat (incomplete).
CC CC REGION 35 1050 Catalytic (approximate).
CC CC REGION 1126 1455 2,4 A, 1 C and 1 AC repeats.
CC CC REGION 1126 1455 Glucan-binding (approximate).
CC CC VARIANT 21 21 V -> I (in strain GS-5).
CC CC VARIANT 81 81 P -> L (in strain MT4239).
CC CC VARIANT 106 106 D -> V (in strain GS-5).
CC CC VARIANT 116 116 S -> A (in strain GS-5 and strain
CC CC MT4467).
CC CC VARIANT 126 126 A -> T (in strain GS-5).
CC CC VARIANT 150 151 SR -> PK (in strain GS-5, strain MT4239
CC CC and strain MT4467).
CC CC VARIANT 256 256 A -> V (in strain GS-5 and strain
CC CC MT4467).
CC CC VARIANT 425 425 R -> N (in strain MT4251).
CC CC VARIANT 519 519 Y -> D (in strain MT4245 and strain
CC CC MT4251).
CC CC VARIANT 538 538 R -> K (in strain MT4245 and strain
CC CC MT4251).
CC CC VARIANT 545 545 Y -> F (in strain MT4245 and strain
CC CC MT4251).
CC CC VARIANT 597 597 N -> D (in strain MT4245, strain MT4251,
CC CC strain MT4467 and strain MT8148).
CC CC VARIANT 600 600 R -> K (in strain MT4245, strain MT4251,
CC CC strain MT4467 and strain MT8148).
CC CC VARIANT 601 601 A -> T (in strain GS-5).
CC CC VARIANT 614 614 M -> T (in strain GS-5).
CC CC VARIANT 727 727 T -> I (in strain MT8148).
CC CC VARIANT 734 734 A -> V (in strain MT8148).
CC CC VARIANT 964 964 L -> F (in strain MT4239).
CC CC VARIANT 1113 1113 N -> Y (in strain MT4239).
CC CC VARIANT 1118 1118 A -> T (in strain MT4239).
CC CC VARIANT 1204 1204 I -> V (in strain GS-5, strain MT4239,
CC CC strain MT4467 and strain MT8148).
CC CC VARIANT 1208 1208 V -> I (in strain MT8148).
CC CC VARIANT 1292 1294 DGH -> NGY (in strain GS-5, strain MT4467

```

```

FT VARIANT 1305 1369 and strain MT8148).
FT VARIANT 1326 1326 Missing (in strain MT4245).
FT VARIANT 1331 1331 I -> V (in strain MT8148).
FT VARIANT 1377 1377 T -> A (in strain GS-5, strain MT4239,
FT VARIANT 1398 1398 strain MT4467 and strain MT8148).
FT VARIANT 1424 1424 R -> K (in strain MT8148).
FT VARIANT 1439 1439 V -> I (in strain MT8148).
FT VARIANT 1444 1444 D -> N (in strain MT4239).
FT VARIANT 1455 1455 V -> I (in strain MT4239 and strain
FT VARIANT 1455 1455 MT8148).
FT CONFLICT 1337 1455 S -> P (in strain MT8148).
FT CONFLICT 1337 1455 ORLYFKSNGVQAKGLITERKGRKIKYVDPNSGVNRRNVVR
FT CONFLICT 1337 1455 TSSGWNWYFGNDGFGSSAVRFRHSRNGVDFDNGYVASHD
FT CONFLICT 1337 1455 QRNHWDYDRDQFGRSSAVRFRHSRNGVDFDNGYVASHD
FT CONFLICT 1337 1455 HASILSLMVFRLRSLQSVKVSNTMILIPMKFVIM
FT CONFLICT 1337 1455 (in Ref. 1).
SQ SEQUENCE 1455 AA; 162966 MW; 3CB455A99A4FEC86 CRC64;

Query Match 96.7%; Score 6905; DB 1; Length 1455;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1323; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

QY 1 MEKVRFKLRKVKRWVTSIASAVVTLTSLSGSLVKADSTDDROQAVTESQASLVTISE 60
DB 1 MEKVRFKLRKVKRWVTSIASAVVTLTSLSGSLVKADSTDDROQAVTESQASLVTISE 60
QY 61 AAKETLTATDTSTATSQPTATVTDNVSTTQSTNTTANTANFVVKPTTSEQAKTDN 120
DB 61 AAKETLTATDTSTATSQPTATVTDNVSTTQSTNTTANTANFVVKPTTSEQAKTDN 120
QY 121 SDKIITTSKAVNRLTATCFKVPANNNTAHPKTVTDKIPIKPKIGKLGKQPSSLSQDDIAA 180
DB 121 SDKIITTSKAVNRLTATCFKVPANNNTAHPKTVTDKIPIKPKIGKLGKQPSSLSQDDIAA 180
QY 181 LGNVKNIRKVGKYYKEDGTLQKNYALNTNGKTFEDETALSNNTLPSKKGNTTND 240
DB 181 LGNVKNIRKVGKYYKEDGTLQKNYALNTNGKTFEDETALSNNTLPSKKGNTTND 240
QY 241 NTNSPAQYNQVYSTDVANFEHVDHYLTAEVWPVKYILKDGKWTQTSTKDFRPLMTWM 300
DB 241 NTNSPAQYNQVYSTDVANFEHVDHYLTAEVWPVKYILKDGKWTQTSTKDFRPLMTWM 300
QY 301 PDQETQROYVNMNAQLGIHOTYNTATSPLOLNLAQIOTKIEBKITAENKTNWLRQTI 360
DB 301 PDQETQROYVNMNAQLGIHOTYNTATSPLOLNLAQIOTKIEBKITAENKTNWLRQTI 360
QY 361 SAFVKTQSAWNSDSEKPPDDHLQKALLYSNNSKLTQANSNYRILNRTPTNOTGKCDPR 420
DB 361 SAFVKTQSAWNSDSEKPPDDHLQKALLYSNNSKLTQANSNYRILNRTPTNOTGKCDPR 420
QY 421 YTADRTIGGYEFLANDVDSNPVQAEQLNWLHFLMNFNGNIYANDPDANFDSIRDAVD 480
DB 421 YTADRTIGGYEFLANDVDSNPVQAEQLNWLHFLMNFNGNIYANDPDANFDSIRDAVD 480
QY 481 NVADALLOIAGDYLKAAGIKHNKDKAANDHLSILEAWSYNDTPYLHDDGDNNMNDNRLR 540
DB 481 NVADALLOIAGDYLKAAGIKHNKDKAANDHLSILEAWSYNDTPYLHDDGDNNMNDNRLR 540
QY 541 LSLYSLAKPLNQSRGMPLTNSLVNRTDQNAETAAPSYSFIRAHDSQVODLIRNIIR 600
DB 541 LSLYSLAKPLNQSRGMPLTNSLVNRTDQNAETAAPSYSFIRAHDSQVODLIRNIIR 600
QY 601 TEINPENVVGYSTFTBEIKKAFEIYNKOLLATEKKYTHYNTALSYALLLTNKSQVPRVYVG 660
DB 601 TEINPENVVGYSTFTBEIKKAFEIYNKOLLATEKKYTHYNTALSYALLLTNKSQVPRVYVG 660
QY 661 DMFTDDGGYMAHKTINYEAITLLKARIKYVSGGAMRNQGVNSSEIITSVRYGKALKA 720
DB 661 DMFTDDGGYMAHKTINYEAITLLKARIKYVSGGAMRNQGVNSSEIITSVRYGKALKA 720
QY 721 TDTGRTTTSQVAVIEGNNPSRLKASDRVVVNNGAHKNQAYRPLLLTTDNGIKAYHS 780
DB 721 TDTGRTTTSQVAVIEGNNPSRLKASDRVVVNNGAHKNQAYRPLLLTTDNGIKAYHS 780

```

QY 781 DQBAAGLVRYNDRGELIFTAADIKGYANPQVSGYLGWVVPVGAADQDVRVAAPST 840
 DB 781 DQBAAGLVRYNDRGELIFTAADIKGYANPQVSGYLGWVVPVGAADQDVRVAAPST 840
 QY 841 DGKSVHONALDSRVNFEFGSNFOATKKEEYTNVIAKNVDKFAEWGTFDEMAPOV 900
 DB 841 DGKSVHONALDSRVNFEFGSNFOATKKEEYTNVIAKNVDKFAEWGTFDEMAPOV 900
 QY 901 SSTDSGFLDSVIONGVAFTDRYDLGISKPNKYGTADDLVAIKALHSKGIKWADWVPDQ 960
 DB 901 SSTDSGFLDSVIONGVAFTDRYDLGISKPNKYGTADDLVAIKALHSKGIKWADWVPDQ 960
 QY 961 MYALPEKEVVTATRVVDKYGTPVAGSQIKNTLYVVDGKSGKQQAQYGGAFLEELQAKYP 1020
 DB 961 MYALPEKEVVTATRVVDKYGTPVAGSQIKNTLYVVDGKSGKQQAQYGGAFLEELQAKYP 1020
 QY 1021 ELFPARKQISTGVPMDSVKIKOWSAKYFNGTILGRGAGYVLKQATNTYFSLVSDNTEFL 1080
 DB 1021 ELFPARKQISTGVPMDSVKIKOWSAKYFNGTILGRGAGYVLKQATNTYFSLVSDNTEFL 1080
 QY 1081 PKSLVNPNGHTSSVTLGVLFDGKGYVYVYSTSGNQAKNAFISLGNWYFDNNGYVMTGAQ 1140
 DB 1081 PKSLVNPNGHTSSVTLGVLFDGKGYVYVYSTSGNQAKNAFISLGNWYFDNNGYVMTGAQ 1140
 QY 1141 SINGANYFSLNGIQLRNAIYDNGNKVLSYNGDGRYENGYYLFGQWRVYFQNGIMAYG 1200
 DB 1141 SINGANYFSLNGIQLRNAIYDNGNKVLSYNGDGRYENGYYLFGQWRVYFQNGIMAYG 1200
 QY 1201 LTRVHGAOVYFDASGFOAQGOITTTADGKLRFDRDSGNQISNRFVNRSGKGEWFLFDHNG 1260
 DB 1201 LTRVHGAOVYFDASGFOAQGOITTTADGKLRFDRDSGNQISNRFVNRSGKGEWFLFDHNG 1260
 QY 1261 VAVTGTVTENGORLFPKPNVQAKGEFIRDANGLYRLYDPNSGNEVRNRFVNRSGKGEWFL 1320
 DB 1261 VAVTGTVTENGORLFPKPNVQAKGEFIRDANGLYRLYDPNSGNEVRNRFVNRSGKGEWFL 1320
 QY 1321 FDHNGIATVGARVAVNG 1336
 DB 1321 FDHNGIATVGTRVAVNG 1336

RESULT 2

GTFFB_STRMU STANDARD; PRT: 1476 AA.
 AC P08987; O69381; O69384; O69387; O69390; O69396;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-MAY-2005 (Rel. 47, Last annotation update)
 DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
 DE (Sucrose 6-glucosyltransferase)
 GN Name=gtfB; OrderedLocusName=SMU.1004;
 OS Streptococcus mutans.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1309;
 RN [1]_NUCLEOTIDE SEQUENCE.
 RP STRAIN=GS-5;
 RC MEDLINE=87308013; PubMed=3040685;
 RX Shiroza T., Ueda S., Kuramitsu H.K.;
 RA "Sequence analysis of the gtfB gene from Streptococcus mutans."; J. Bacteriol. 169:4263-4270(1987).
 RL [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=MT4239 / Serotype c, MT4245 / Serotype e, MT4251 / Serotype f,
 RC MT4467 / Serotype e, and MT8148 / Serotype c;
 RX MEDLINE=98231643; PubMed=9570124;
 RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
 RA Kimura S., Hamada S.;
 RT "Molecular analyses of glucosyltransferase genes among strains of Streptococcus mutans."; FEMS Microbiol. Lett. 161:331-336(1998).
 RL FEMS Microbiol. Lett. 161:331-336(1998).

RN [3] NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RP STRAIN=UAI59 / ATCC 700610 / Serotype c;
 RC MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;
 RX Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
 RA Carlson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
 RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,
 RA Ferretti J.J.;
 RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental pathogen."; Acad. Sci. U.S.A. 99:14434-14439(2002).
 RL Proc. Natl.
 CC -!- FUNCTION: Production of extracellular glucans, that are thought to play a key role in the development of the dental plaque because of their ability to adhere to smooth surfaces and mediate the aggregation of bacterial cells and food debris.
 CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-fructose + (1,6-alpha-D-glucosyl) (n+1).
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both forms of glucans.
 CC -!- SIMILARITY: Belongs to the glucosyl hydrolase 70 family.
 CC -!- SIMILARITY: Contains 10 cell wall binding repeats.
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.
 CC EMBL; M17361; AAA88588.1; -; Genomic DNA.
 DR EMBL; D88651; BAA26101.1; -; Genomic DNA.
 DR EMBL; D88654; BAA26105.1; -; Genomic DNA.
 DR EMBL; D88657; BAA26109.1; -; Genomic DNA.
 DR EMBL; D88660; BAA26113.1; -; Genomic DNA.
 DR EMBL; D89977; BAA26119.1; -; Genomic DNA.
 DR EMBL; AE014940; AAN58705.1; -; Genomic DNA.
 DR PIR; B33135; B33135.
 DR HSSP; P06653; 1H8G.
 DR InterPro; IPR002479; Cell_wall_bd_put.
 DR InterPro; IPR003318; Glyco_hydro_70.
 DR Pfam; PF01473; CW_binding_1; 4.
 DR Pfam; PF02324; Glyco_hydro_70; 1.
 KW Complete proteome; Dental Caries; Glycosyltransferase; Repeat; Signal;
 KW Transferase.
 FT SIGNAL 1 34 Potential.
 FT CHAIN 35 1476 Glucosyltransferase-I.
 FT REPEAT 1097 1130 A repeat.
 FT REPEAT 1161 1210 1.
 FT REPEAT 1225 1275 2.
 FT REPEAT 1290 1340 3.
 FT REPEAT 1355 1405 4.
 FT REPEAT 1420 1470 5.
 FT REGION 35 1051 Catalytic (approximate).
 FT REGION 1097 1476 Glucan-binding (approximate).
 FT REGION 1161 1470 5 X tandem repeats.
 FT VARIANT 62 62 S -> T (in strain MT4239).
 FT VARIANT 65 65 T -> I (in strain GS-5).
 FT VARIANT 68 68 V -> A (in strain GS-5, strain MT4245, strain MT4251, strain MT4467 and strain MT8148).
 FT VARIANT 78 78 Q -> P (in strain MT4251).
 FT VARIANT 86 86 I -> S (in strain GS-5, strain MT4245, strain MT4251, strain MT4467 and strain MT8148).
 FT VARIANT 89 89 S -> F (in strain MT4251).
 FT VARIANT 168 168 K -> N (in strain MT4251).
 FT VARIANT 276 276 S -> D (in strain GS-5, strain MT4467 and strain MT8148).
 FT VARIANT 399 399 N -> R (in strain MT4239).
 FT VARIANT 474 474 I -> T (in strain MT4239).
 FT VARIANT 512 512 K -> R (in strain MT8148).

FT	VARIANT	519	519	F -> Y (in strain MT8148).
FT	VARIANT	701	701	T -> I (in strain MT8148).
FT	VARIANT	708	708	A -> V (in strain MT8148).
FT	VARIANT	938	938	F -> L (in strain MT8148).
FT	VARIANT	952	957	FGKPE -> YGTPVA (in strain GS-5, strain MT4239 and strain MT4467).
FT	VARIANT	963	964	SV -> NT (in strain GS-5, strain MT4239 and strain MT4467).
FT	VARIANT	968	970	ADS -> VDG (in strain GS-5, strain MT4239 and strain MT4467).
FT	VARIANT	1086	1086	A -> T (in strain MT4239).
FT	VARIANT	1158	1158	S -> N (in strain MT4239).
FT	VARIANT	1163	1163	H -> Y (in strain MT4251).
FT	VARIANT	1168	1168	E -> K (in strain MT8148).
FT	VARIANT	1182	1182	Y -> C (in strain MT8148).
FT	VARIANT	1234	1234	A -> P (in strain MT4239).
FT	VARIANT	1263	1263	R -> H (in strain GS-5 and strain MT4467).
FT	VARIANT	1263	1263	R -> P (in strain MT8148).
FT	VARIANT	1264	1264	Y -> H (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).
FT	VARIANT	1272	1272	S -> G (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).
FT	VARIANT	1329	1329	H -> Y (in strain GS-5 and strain MT4467).
FT	VARIANT	1394	1394	Y -> H (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).
FT	VARIANT	1402	1402	S -> G (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).
FT	VARIANT	1459	1459	Y -> H (in strain MT4467).
FT	CONFLICT	570	570	R -> A (in Ref. 1).
FT	CONFLICT	800	817	AQDVRVAASTAPSTDGK -> LKMFALRLARPHQOMA (in Ref. 1).
FT	CONFLICT	1310	1310	H -> L (in Ref. 1).
SQ	SEQUENCE	1476	AA; 165847	MM; 9C6E09F731B4CBFC CRC64;
Query Match				
Best Local Similarity 74.9%; Score 5350; DB 1; Length 1476;				
Matches 1018; Conservative 94; Mismatches 168; Indels 44; Gaps 11;				
QY	1	MEKKVRFKLKVKKRWVTVSIA	SVVTLTSLGSLVKADSTDDROQAVTESQASLVTTSE	60
DB	1	MDKKVRYKLRKVKRWVTVS	ASAVMTLTLSGLGLKADSNESKQISNDSTVTVANE	60
QY	61	AAKETLTATDTATSTATSQPTATVTDNVSTNTQSTNTTANTANFVVKPTTTSBQAKTDN	120	
DB	61	--ESNVTTTQVTSQEAASQTNHTV----	TTISSSTSVNPKVSNPYTVGETA--SN	111
QY	121	SDKIITTSKAVNRLTATGKFPANNNTAHPKTVTDKIVPIKIGKIKQPSSLSODDIAA	180	
DB	112	GEKLNQTTTVDKTSSEA-----	AANNISKQTTTEAD-----TDVIDDSNAA	151
QY	181	----LGNVKNIRKVNAGKYYYKEDGTLQKNYALNNGKTFPPDETGALSNNLTLPKKGNI	236	
DB	152	NIQILEKLPNVKEIDGKYYYDNNGKVRNFTLIADGKILHFDGTGAYTDTSDTVNKDI	211	
QY	237	TNNDNTNSFAQNVYSTDVANFEHVDHYLTAESWYRPKIYKDKGTWTQSTEKDPRPLL	296	
DB	212	VTT-RSNLYKYNQVYDRSAQFEHVDHYLTAESWYRPKIYKDKGTWTQSTEKDPRPLL	270	
QY	297	MTWMPDQETOROVYNNMNAQLGIHOTYNTATSPLOLNLAQTIOTKIEKTAENKNWL	356	
DB	271	MTWMPSEQTOROVYNNMNAQLGINKTYDDTSLNQLQNLNAAATQAKIEAKITTLKNKTDL	330	
QY	357	RTQISAFVKTSQAMNSDSKPPDDHLQKGLALYSNNKSLTQANSNYRILNRPNTQTKG	416	
DB	331	RTQISAFVKTSQAMNSDSKPPDDHLQNGAVLYDNEGKLTTPYANSNYRILNRPNTQTKG	390	
QY	417	KDPRYTADRTIGGYEFLANDVDNSNPVQABQLNLWFLMFGNIYANDPDANFDSIRV	476	
DB	391	KDPRYTADRTIGGYEFLANDVDNSNPVQABQLNLWFLMFGNIYANDPDANFDSIRV	450	
QY	477	DAVDNVDADLLQIAGDYLKAAGIHKNDKAANDHLSILEAWSYNDTPYVLHDDGDNMINWD	536	

DB	451	DAVDNVDADLLQIAGDYLKAAGIHKNDKAANDHLSILEAWSYNDTPYVLHDDGDNMINWD	510	
QY	537	NRLRLSLLYSLAKPLNQSGMNPILITNSLVNRTDDNAETAAPVPSYSFIRAHDSYVODLIR	596	
DB	511	NKLRSLLLSLAKPLNQSGMNPILITNSLVNRTDDNAETAAPVPSYSFIRAHDSYVODLIR	570	
QY	597	NIIRTEINPNVGVSTFTEEBIKKAPFIYNKDLATEKKYTHYNTALSVALLLTNKSSVPR	656	
DB	571	DIIKABINPNVGVSTFTEEBIKKAPFIYNKDLATEKKYTHYNTALSVALLLTNKSSVPR	630	
QY	657	VYVGMFTDDGOYMAHKTINTEAETLLKARIKYVSGOAMRNQOVGNSEIITSVRYGKG	716	
DB	631	VYVGMFTDDGOYMAHKTINTEAETLLKARIKYVSGOAMRNQOVGNSEIITSVRYGKG	690	
QY	717	ALKATDTGDRTRTSGVAVIEGNPNPSRLKASDRVVNMGAAHKNQAVRPLLLTTDNGIK	776	
DB	691	ALKATDTGDRTRTSGVAVIEGNPNPSRLKASDRVVNMGAAHKNQAVRPLLLTTDNGIK	750	
QY	777	AYHSDQEAAGLVRYTNDRGELIFTAADIKGYANPOVSGYLGWVPVGAADODVRAAST	836	
DB	751	AYHSDQEAAGLVRYTNDRGELIFTAADIKGYANPOVSGYLGWVPVGAADODVRAAST	810	
QY	837	APSTDGKSVHQNAAALDSRVMEFGFSNFQAFATKKEEYTNVVIKXNDVKFAEWGVTDFEMA	896	
DB	811	APSTDGKSVHQNAAALDSRVMEFGFSNFQAFATKKEEYTNVVIKXNDVKFAEWGVTDFEMA	870	
QY	897	POYVSSSTDGSLDSVIONGYAFTDRYDILGISKPNKYGTADDLVKAIKALHSGIKVMADW	956	
DB	871	POYVSSSTDGSLDSVIONGYAFTDRYDILGISKPNKYGTADDLVKAIKALHSGIKVMADW	930	
QY	957	VPDQMYALPEKEVVTATRVKGYTPVAGSOIKNTLYVVDGKSSGKQQAQYGGAFLEELQ	1016	
DB	931	VPDQMYAFPEKEVVTATRVKFGKPEVGSQKSVLYVADSKSSGKQQAQYGGAFLEELQ	990	
QY	1017	AKYPPELPARKOISTGVPMDPSPVKIKOWSAKYNFNGTNIILGRGAGYVLKQOATNTYFSLVSD	1076	
DB	991	AKYPPELPARKOISTGVPMDPSPVKIKOWSAKYNFNGTNIILGRGAGYVLKQOATNTYFN-1SD	1049	
QY	1077	N---TFLPKSLVNPNHGTSSTVTGLVFDGKGYVYVYSTSGNQAQNAFISLGNWNYFDNNG	1133	
DB	1050	NKEINFLPKTLN-----QDSQVGFSDGKGYVYVYSTSGYQAKNTFISEGDKWYFDNNG	1104	
QY	1134	YMTVGAQSGINGANYFFLSNGIQLRNAIYDNGKVLVSYGNDGRRYENGYYLF-QQWRYF	1192	
DB	1105	YMTVGAQSGINGANYFFLSNGIQLRNAIYDNGKVLVSYGNDGRRYENGYYQPMGVRHFF	1164	
QY	1193	QNGIMAVGLTRVHGAVQYFDASGFOAKGQFTTTADGKLYRFDSDSGNQISNRFVRSKGE	1252	
DB	1165	NNGEMSVGLTVIDGOVQYFDEMGYQAKGFVTADGKLYRFDKQSGNMYRNFRIENEGK	1224	
QY	1253	WFLFDHNGVAVTGTVTTFNGQRLYKPNQVQAKGEFIRDANGYLRYDDPNSGNEVNRNFR	1312	
DB	1225	WLYLGEDAAVTGTSQTINGQHLVFRANGVQKGEFVTDRYGRISYIYDSNSGDIQRNFR	1284	
QY	1313	NSKGEWFLFDHNGTAVTCARVUNG	1336	
DB	1285	NAQQQWFYFDNNGYAVTGARTING	1308	
RESULT 3				
Q59983 STRE PRELIMINARY; PRT; 1590 AA.				
ID	Q59983	STRE	PRELIMINARY;	PRT; 1590 AA.
AC	Q59983			
DT	01-NOV-1996	(TRENBLrel. 01, Created)		
DT	01-NOV-1996	(TRENBLrel. 01, Last sequence update)		
DT	01-MAR-2004	(TRENBLrel. 26, Last annotation update)		
DE	Glucosyltransferase-I precursor (EC 2.4.1.5).			
GN	Name=gcf1;			
OS	Streptococcus sobrinus.			
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;			
OC	Streptococcus.			
OX	NCBI_TaxID=1310;			

```

RN  [1]
RP  NUCLEOTIDE SEQUENCE.
RC  STRAIN=OMZ176;
RX  MEDLINE=94146405; PubMed=8312602;
RA  Sato S., Inoue M., Handa N., Aizawa Y., Isebe Y., Katayama T.;
RT  "DNA sequence of the glucosyltransferase gene of serotype d
RL  Streptococcus sobrinus";
RL  DNA Seq. 4:19-27(1993).
RP  [2]
RN  PROTEIN SEQUENCE.
RX  MEDLINE=9122498; PubMed=1827439;
RA  Mooser G., Hefsa S.A., Paxton R.J., Shively J.E., Lee T.D.;
RT  "Isolation and sequence of an active-site peptide containing a
RT  catalytic aspartic acid from two Streptococcus sobrinus alpha-
RT  glucosyltransferases.";
RL  J. Biol. Chem. 266:8916-8922(1991).
DR  EMBL; DI3858; BAA02976.1; -; Genomic_DNA.
DR  PIR; A39841; A39841.
DR  HSSP; P06653; LHGX.
DR  GO; GO:0047849; F:dextranucrase activity; IEA.
DR  GO; GO:0016757; F:transferase activity, transferring glycosyl. .; IEA.
DR  GO; GO:0009250; P:glucan biosynthesis; IEA.
DR  InterPro; IPR002479; CW binding.
DR  InterPro; IPR003318; Glyco_hydro_70.
DR  Pfam; PF01473; CW_binding_1; 2.
DR  Pfam; PF02324; Glyco_hydro_70; 1.
KW  Glucosyltransferase; Signal; Transferase.
FT  SIGNAL 1 38 Potential.
FT  CHAIN 39 1590 glucosyltransferase-I.
SQ  SEQUENCE 1590 AA; 175956 MW; C3C83A57CF3C2B0E CRC64;

Query Match 59.0%; Score 4214.5; DB 2; Length 1590;
Best Local Similarity 60.3%; Pred. No. 1.6e-192;
Matches 811; Conservative 166; Mismatches 296; Indels 53; Gaps 14;

QY 1 MEKVRFLKRVKRWVTSIASAVVTLISLSGLVKAD---STDRQAVTESQASLVT 57
DB 1 MEKVRFLKRVKRWVTSIASAVVTLISLSGLVKAD---STDRQAVTESQASLVT 57
QY 58 TSEAAKETLATDTSTATSOPATVTDVNTSTNQNTTANTANFVVKPTTSEQAK 117
DB 59 NQATDQTSIA---ATATSEQASTDAATDQSAEQTGGTAST--DTAAQTITNANE 113
QY 118 TDNSDKIITTSKAVNRLTATGKFPVANNNTAHPKVTDKIPIKIGLKOPSSLSQDD 177
DB 114 -----KWPTEENEN---QGFTDEM-----LAEAKNVATAESDS 143
QY 178 IAA-IGNVKNIRKNGKYYKEDGTLQKQYALNNGTKFTFDETCALSNNT-LPSKGN 235
DB 144 IPSDLAKSNVRQVDGKYYQDQGNVKNFVSVGDKIYYFDETGAYKDTSKVDKSS 203
QY 236 ITNNNTNSFAQYNQVYSTDVANFEHVDHYLTAEWSYRKYILKQKWTQSTKDRPL 295
DB 204 SAVSQNATIFAANNRAYSTSAKNFEAVDNYLTADSWYRPKSILKQKWTQSTKDRPL 263
QY 296 LMTWPDQETQRYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 355
DB 264 LMAWPDQETQRYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 323
QY 356 LQQTSAFVKTSQANSSEKPFDPHQLQKALLYNNKSLTSQANSYRILNRTPTNQG 415
DB 324 LREAIASFVKTPQWNGSEKPYDHLQNGALLFQNTDLTPTQSNRYLLNRTPTNQG 383
QY 416 KDPRTY--ADRTIGYFELLANDVNSNPVQVQAEQLNWLHFLMFGNYANDPDNFS 473
DB 384 SLDSFTYVNDPDLGGYFELLANDVNSNPVQVQAEQLNWLHFLMFGNYANDPDNFS 443
QY 474 IRVDAVDNVDADLLQIAGYLKAAKGIHKNDKAAANDHLSILEAWSYNDTPYLHDDGDNNI 533
DB 444 IRVDAVDNVDADLLQIAGYLKAAKGIHKNDKAAANDHLSILEAWSYNDTPYLHDDGDNNI 503
QY 534 NMDNRLRLSLYLAKPLNQRSGMNPPLITNSLVNRTDQNAETAAPVPSYFIRAHDSVQD 593
DB 534 NMDNRLRLSLYLAKPLNQRSGMNPPLITNSLVNRTDQNAETAAPVPSYFIRAHDSVQD 593

```

```

DB 504 NMDNRLRLSLYLAKPLNQRSGMNPPLITNSLVNRTDQNAETAAPVPSYFIRAHDSVQD 563
QY 594 LIRNIIRTEINPNVGVYSTTEIEIKKAFIYNKDLATEKKYTHYNTALSYALLLNKSS 653
DB 564 IIRDIKABINPNSFGYSFTQIEIQAFIYNEDLKKTKYTHYVNPVPSYLLLNKGS 623
QY 654 VPRVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 713
DB 624 IPRVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 683
QY 714 GKGALKATDGTGRTTSTGAVIEGNNPSRLKASDRVNVVVMGAHKNQAYRPLLLTTDN 773
DB 684 GKGALKQSDKGDATRTTSGVVMGNQPNFSLDGK-VVALNMGAAHANQSYRALMVSTKD 742
QY 774 GIKAYHSDQEA--AGLVRYTNDRGELIFTAADIKGVANPQVSGYLCVWVPVGAADQDVR 831
DB 743 GVATYATDADASKAGLVKRTDENGILYFLNDDLKGVANPQVSGFLOVWVPVGAADQDIR 802
QY 832 VAASTAPSTDGKSHVQNAALDSRVMPGFSNFQAFATKKEEYTNVVIKAKVDFAEWGV 891
DB 803 VAASTDASTDGKSLHQDAAMDSCRVMFEGFSNFQSPATKEEYTNVVIKAKVDFAEWGV 862
QY 892 DFEMAPQVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 951
DB 863 DFEMAPQVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 922
QY 952 VNADWVPDQMYALPEKEVVTATRVKGYTPVAGSQIKNTLYVVDGKSGKQDQAKYGGAF 1011
DB 923 VNADWVPDQMYALPEKEVVTATRVKGYTPVAGSQIKNTLYVVDGKSGKQDQAKYGGAF 982
QY 1012 LLEOAKYPELFPARKQISTGVPMDFPSVKIKOWSAKYFNGTILGRGAGYVLKQDQATNTYF 1071
DB 983 LDELKKEYPELFTKQISTGQAIQDPSVKIKOWSAKYFNGTILGRGAGYVLKQDQATNTYF 1042
QY 1072 SILVSNTEFLPKSLVNPNGHTSSVTGLVDPGKGYVYVYST-SGNQAKNAFISLGNWYVYD 1130
DB 1043 NVASDTLFLPSSLLG-----KVVEBSGIRYDGRGYIYNSSATGDQVKASPIEAGNLYYFG 1097
QY 1131 NNGYVMTGAQSNINGANYFLSNGIOLRNAIYDNGKNVLSYVYVYVYVYVYVYVYVYVYVYVYVY 1190
DB 1098 KQGYVMTGAQINGANYFLSNGIOLRNAIYDNGKNVLSYVYVYVYVYVYVYVYVYVYVYVYVY 1157
QY 1191 YFQNGIMAVGLTRVHGVQYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 1250
DB 1158 YFQNGIMAVGLTRVHGVQYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 1217
QY 1251 GEFWFLPDHNGVAVTGTVTGNGQRLYFKPVGQAKGEFIRDANGYLYRYDYPNSGNEVNRFP 1310
DB 1218 GHWYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 1277
QY 1311 VRNSKGEWFLPDHNGIYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 1336
DB 1278 IEDKAGNWFYLGKOGAAVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 1303

```

RESULT 4

```

Q55263 9STRE
ID Q55263 9STRE PRELIMINARY; PRT; 1590 AA.
AC Q55263;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE GTF-1.
GN Name=Glucosyltransferase;
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1310;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 33478;
RA Sato S.;
RT "DNA and amino-acid sequences of water-insoluble-glucan synthetase

```

RT produced from Streptococcus sobrinus ATCC 33478. ";
 RL Kagoshima Daigaku Shigakubu Kiyo 16:23-29(1996).
 RN [2]

PROTEIN SEQUENCE.

RA MEDLINE=91224988; PubMed=1827439;
 RX Moser G., Hefta S.A., Paxton R.J., Shively J.E., Lee T.D.;
 RT "Isolation and sequence of an active-site peptide containing a
 RT catalytic aspartic acid from two Streptococcus sobrinus alpha-
 RT glucosyltransferases."; J. Biol. Chem. 266:8916-8922(1991).
 RL J. Biol. Chem. 266:8916-8922(1991).
 DR EMBL; D63570; BAA09792.1; -; Genomic_DNA.
 DR PIR; A39841; A39841.
 DR HSP; P06653; LGVM.
 DR GO; GO:0009250; P-glucan biosynthesis; IEA.
 DR InterPro; IPR002479; CW_binding.
 DR InterPro; IPR003318; Glyco_hydro_70.
 DR Pfam; PF01473; CW_binding_1; 3.
 DR Pfam; PF02324; Glyco_hydro_70; 1.
 SQ SEQUENCE 1590 AA; 176058 MW; 9DF7A3F2C6B4FD43 CRC64;

Query Match 58.9%; Score 4207.5; DB 2; Length 1590;
 Best Local Similarity 60.2%; Pred. No. 3.4e-192;
 Matches 810; Conservative 185; Mismatches 298; Indels 53; Gaps 14;

QY 1 MEKVRFKLRKVKRWVTVSIAVAVTLTSLGSLVKAD---STDRQQAUTESQASLVT 57
 DB 1 MEKVRFRKVRKRWVTVLUSVATSLALGASVASADTDASDDSNQAVVTGQDT--T 58
 QY 58 TSEAAKETLTATDTSTATSQPTATVDNVTSTNTQSTNTTANTANFVVKPTTSEQAK 117
 DB 59 NQOATDOTSIA---ATATSEQASDADTDQASAAEQTQGTAST-DTAAQTTTANEA- 113
 QY 118 TNSOKIITTSKAVNRLTATGKFPVANNHTAPKVTVDKIVPIKPIKIGLKOPSSLSQDD 177
 DB 114 -----KWPPTENEN---QGFDEM-----LAERKNVATAESDS 143
 QY 178 IAA-LGNVKNIRKNGKYYYKEDGTLOKNYALNNGKTFPDETCALSNNT-LPSKGN 235
 DB 144 FPSDLAKSNVQVQDKYIYYDDGNVKNFVAVSGDKIYYFDETCAYKDTSKVDADKSS 203
 QY 236 ITNNNTNSFAQYNQVYSTVANFEHVDHLYTAESWYRPKYILKDGKTWTQSTKDFRPL 295
 DB 204 SAVSQNATIFAANNRAYSTSAENFEAVDNYLTADSWYRPKSLKDGKTWTESKDDFRPL 263
 QY 296 LMTWPPDQSTQRYNNYMAQLGIGHTYNTATSPQLNLAQTIQTKIEEKTAEKNTW 355
 DB 264 LMAWPPDTEKRYNNYNNLVVIGIDKTYTAETSQADLTAAAEVLQVARIQKITTEQNTKW 323
 QY 356 LRQTISAFVKTSANWSDSEKPPDDHLOKGLLYSNNSKLTQANSNYRILARTPTNOTG 415
 DB 324 LREAISAFVKTPQWNGSEKPYDDHLOKGLKLFQNSDLTPDTQSNYRLLARTPTNOTG 383
 QY 416 KKDPRYT--ADRTIGYBPELLANDVNSNPVQAEQLNWLHFLMFGNTIYANDPDANFDS 473
 DB 384 SLDSRTYNANDPLGGEYBPELLANDVNSNPVQAEQLNWLHLYLNFSGSIYAKDADANFDS 443
 QY 474 IRVDAVNDVADLLQIAGDYLKAAGIHKNDKAANDHLSILEAWSYNDTPYLHDDGNMI 533
 DB 444 IRVDAVNDVADLLQISDYLKAAYGIDKNNKANNHVSIVEAWSNDTPYLHDDGNILM 503
 QY 534 NNDNRLRLSLYLAKPLNQRSGMPLITNSLVNRTDQNAETAAPVPSYFIRAHDSVQD 593
 DB 504 NMDNFRSLWSLAKPLDKRSLGNPLIHNSLVDRVDREVTVPYSYFAAHSEVQD 563
 QY 594 LIRNIRTEINPNVNGYSPTTEIKKAFIYNKDLILATEKKYTHYNTALSYALLLTNKS 653
 DB 564 IIRDIKAEINENSGYSFTQEBIDQAFKIYNEDLKKTKYTHYNPVLSYLLLTNKS 623
 QY 654 VPRVYGDFTDDGQYMAKHTNYEAIETLLKARIKYSGGQAMNQVGNSEIITSVRY 713
 DB 624 IPRVYGDFTDDGQYMAKHTNYEAIETLLKARIKYSGGQAMNQYQIENGSEIITSVRY 683
 QY 714 GKGLKATDGTGRTTTSVAVIEGNNPSLRKASDRVVVNWGAHKQAYRPLLTDTN 773

DB 684 KGKALKQSDKGDATRTTSVGVMGNQNFSLDGK-VVALNWGAHAHQEYRALAVSTKD 742
 QY 774 GIKAYHSDQEA--AGLVRYTNRDRELIFTAADIKIYANPQVSGYLGWVPVCGAADDQVR 831
 DB 743 GVATYATDADASKAGLVRTDENGLYFLNDDLKGVANPQVSGFLQWVPVCGAADDQDTR 802
 QY 832 VAASAPSTDCGKSVHONAAALDSRVMEFGFSNFOAFATKKEEYTVVIAKNVDKFAEWGVT 891
 DB 803 VAASDTASTDCGKSLHQDAMDSRVNFBGFSNFOAFATKKEEYTVVIAKNVDKFAEWGVT 862
 QY 892 DPEMAPQVSVSTGSLFDSVIQNGYAFTRDYLGIISKPNKYGTADDLKAIALKHSKGIK 951
 DB 863 DPEMAPQVSVSTGSLFDSVIQNGYAFTRDYLGIISKPNKYGTADDLKAIALKHSKGIK 922
 QY 952 VVADWVPDQMYALPEKVVATRVKGYTPVAGSQIKNTLYVVDGKSGKQOQAKYGGAF 1011
 DB 923 VVADWVPDQMYTPPKQEVTVTRTDKFGKPIAGSQIHNLSLYVTDTKSSGDDYQAKYGGAF 982
 QY 1012 LEELOAKYPELPARKQIISTGVPMDSVKIKOMSAKYFNGTILGRGAGVYLDQATNTVF 1071
 DB 983 LDELKKEYPELPARKQIISTGVPMDSVKIKOMSAKYFNGTILGRGAGVYLDQATNTVF 1042
 QY 1072 SLVSDNTLPLKSLVNPNGHTSSVTGLVFDGKGYVYVYST-SGNQAKNAFISLGNWYFD 1130
 DB 1043 NVASDTLPLPSLLG-----KWVESGIRYDGKGYIYNSSATGDQVKAFITEAGNLYYFG 1097
 QY 1131 NNGVMTGAQINGANYFLSNGIQLRNAIYDNGHKNVLSYNGDGRYVNGYLLFGQOVR 1190
 DB 1098 KDGVMVTGAQINGANYFLSNGIQLRNAIYDNGHKNVLSYNGDGRYVNGYLLFGQOVR 1157
 QY 1191 YFONGIMAVGLTRVHGAVOYFASFOAKGQFIITADGKLYRFDSDSGNOISNRFVRNSK 1250
 DB 1158 YFONGIMAVGLTRVHGAVOYFASFOAKGQFIITADGKLYRFDSDSGNOISNRFVRNSK 1217
 QY 1251 GEWFLFDHNGVAVTGTVTGNGORLYFKPNGVOAKGEFIRDANGYLYRYPDPSNGEVRNR 1310
 DB 1218 GHWYLLGKDGAVVATGAQTVGKQKLYFEANGEQVKGDFVTSHEGKLYFDVDSGDWMTDTF 1277
 QY 1311 VRNSKGEWFLFDHNGIATGARVNG 1336
 DB 1278 IEDKAGNWFYLGKDGAAVSGAQTIRG 1303

RESULT 5

GTFL STRD
 ID GTFL STRD STANDARD; PRT; 1597 AA.
 AC P11001;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 13-SEP-2005 (Rel. 48, Last annotation update)
 DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
 DE (Sucrose 6-glucosyltransferase).
 GN Name=gtfI;
 OS Streptococcus downei (Streptococcus sobrinus).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 RN NCBI_TaxID=1317;
 RX [1]
 RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
 RC STRAIN=MF28;
 RX MEDLINE=87308014; PubMed=3040686;
 RA Ferretti J.J., Gilpin M.L., Russell R.R.B.;
 RT "Nucleotide sequence of a glucosyltransferase gene from Streptococcus
 RL J. Bacteriol. 169:4271-4278(1987).
 CC -!- FUNCTION: Production of extracellular glucans, that are thought to
 CC play a key role in the development of the dental plaque because of
 CC their ability to adhere to smooth surfaces and mediate the
 CC aggregation of bacterial cells and food debris.
 CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-
 CC fructose + (1,6-alpha-D-glucosyl) (n+1).
 CC -!- SUBCELLULAR LOCATION: Secreted.

CC -|- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes
CC water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
CC forms of glucans.
CC -|- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
CC -|- SIMILARITY: Contains 19 cell wall binding repeats.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL: M17391; AAC63063.1; -; Genomic DNA.
DR InterPro: IPR002479; Cell wall bd put.
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; CW binding 1; 4.
DR Pfam: PF02324; Glyco_hydro_70; 1.
KW Dental caries; Glycosyltransferase; Repeat; Signal; Transferase.
FT SIGNAL 1
FT CHAIN 38 Potential.
FT REPEAT 1099 1132 Glucosyltransferase-1.
FT REPEAT 1163 1213 AC repeat.
FT REPEAT 1229 1277 AC repeat.
FT REPEAT 1292 1342 AC repeat.
FT REPEAT 1352 1399 B repeat.
FT REPEAT 1406 1455 AC repeat.
FT REPEAT 1465 1512 B repeat.
FT REPEAT 1519 1568 AC repeat.
FT REPEAT 1582 1597 A repeat (incomplete).
FT REGION 39 1050 Catalytic (approximate).
FT REGION 1099 1597 1.25 A, 2 B and 5 AC repeats.
FT REGION 1099 1597 Glucan-binding (approximate).
SQ SEQUENCE 1597 AA; 177080 MW; B9E86A200868798E CRC64;

Query Match 58.7%; Score 4194; DB 1; Length 1597;
Best Local Similarity 59.6%; Pred. No. 1.5e-191;
Matches 807; Conservative 186; Mismatches 299; Indels 62; Gaps 14;

QY 1 MEKVRFLKVRKVRWTVSYIASAVVLTLSLGSLLVKAD-----STDROQAVTESQASLV 56
DB 1 MEKNERFQHKVKKRWTVSYIASATMLASALGASVASADTETVSEDSNQAVLTADQTTN 60

QY 57 TTSEAKETLTATDTSTATSQPTATVTDNVSTTNQSTNTANTANFVVKPTTTSEQA 116
DB 61 ODTEQTSVAATATSEQASTDAATQASATDQASAAEQTQGTAST-DTAAQTNTANEA 119

QY 117 K---TDNSDKIT---TSKANRLTATGKVFVPANNNTAHPKVTVDKIVPKIKGLKOP 170
DB 120 KVVPTENENQVTDMLAEAKNVATAESNSIPS----- 152

QY 171 SSLSQDDTAALGNKVNIRKVGKYYKEDGTLQKNYALNNGKTFPFDGTGALSNNTL- 229
DB 153 -----DLAKMSNVK---QVDGKYYYDQDGNKKNFVSVGEKIYYFDGTGAYKDTSKV 203

QY 230 -PSKGNITNNDNTNSFAQYNQVSYTDVANFHVHDLTAESWYRPKYLLKDGKTWOST 288
DB 204 EADKSGSDISKET-TFAANNRAYSTSAENFEADNYLTADSWYRPSKLLKDGKTWESS 262

QY 289 EKDFPRLMTWPDQETQRYNNYNAQLGHIHQNTNTATSPQLNLAAQTITQKIEKIT 348
DB 263 KDDFRPFLMAWPDFTETKRYNNYNNKVVGDITKYTAETSQADLTAAAEVLQARIEKIT 322

QY 349 AEKNTNWLRLQTSASFVKTSQANNSSEKFPDHLQKALLYSNNKLTISQANSYRLNR 408
DB 323 TEQNTKWLREALISAFVKTPQWNGSEKPYDHLQNGALKFNQSDLTPTDTSYRLNLR 382

QY 409 TPTNQTGKDPRT--ADRTIGGYEFLANDVNSNPVQARQLNWLHFLMFGNIYAND 466
DB 393 TPTNQTGSLDSFTYNDANPLGGYELLALLANDVNSNPVQARQLNWLHFLMFGNIYAKD 442

QY 467 PDANFDSIRVDVNDVADLLQIAGDYLKAAKGIIHNKDKAANDHLSILEAWSNDTPYLH 526
|||||

DB 443 ADANFDSIRVDVNDVADLLQIASSDYLKAAAYGIDKNNKNNANNHVSIVEAWSNDTPYLH 502
QY |||||
DB 527 DDCDNMINMDNRLRLSLLYSLAKPLNORSGMPLITNSLVNRRTDDNAETAAPSYSFIRA 586
DB 503 DGDGDLMMNDKFRSLWSLAKPLDKRSLGPLNLSLVDREVDDREVETVSYSPARA 562
QY HDBSEVQDLIRNIRTEINPNVVGYSFTTBETBEIKAFIYNNKDLATEKKYTHYNTALSAL 646
DB 563 HDBSEVQDLIRDIKAIKAINPNAFGYSFTQBDIDQAFKIYNEDLKKTKYTHYNVPLSYTL 622
QY 647 LITNKSVPVRYVGMFTDDGQYMAHKTINYEAIETLLKARIKYVSGGQMRNQOVGNSE 706
DB 623 LUTNKSIPRVYVGMFTDDGQYMAHKTINYEAIETLLKARIKYVSGGQMRNQOVGNSE 682
QY 707 IITSVRYGKALKATDGTGRTTSTGVAIVIEGNPSPRLKASDRVVVNNGAHKNQAYRP 766
DB 683 IITSVRYGKALKQSDKGDATTTSGVGVVGMQNPFSLDGK-VVALNMGAAHANQYRA 741
QY 767 LLTTDNGIKAYHSDQEA--AGLVRYTNDRGLIFTAADIKGYANPQVSGYLGWVPVGA 824
DB 742 LMVSTKQGVATYATDADASKAGLVKRTDENGVLFLNDLKGANPQVSGFLQVWVPVGA 801
QY 825 AADQVRVAASTAPSTDGKSVHONAAALDSRVMPFEGSFQAFATKEEYTNVVIANKVDK 884
DB 802 ADDQDIRVAASDTASTDGKSLHODAAAMDSRVMPFEGSFQAFATKEEYTNVVIANNVDK 861
QY 885 FAEWGVTDFEMAPQVYVSSSTDGSLFSLDSVIONGVAFYDRLGTSKPNKYGTADDLVKAIKA 944
DB 862 FVSWGITDFEMAPQVYVSSSTDGSLFSLDSVIONGVAFYDRLGTSKPNKYGTADDLVKAIKA 921
QY 945 LHSKGIKVMADWVPQYVVALPEKVVYVTRVDKYGTPVAGSQIKNTLYVVDGKSSGKDOQ 1004
DB 922 LHAQGLKVMADWVPQYVVALPEKVVYVTRVDKYGTPVAGSQIKNTLYVVDGKSSGKDOQ 981
QY 1005 AKYGAFLEELQAKYPELFAKQISTGVPMDSVKIKQWSAKYFNGTNI LGRGAGVTLKD 1064
DB 982 AKYGAFLEELQAKYPELFAKQISTGVPMDSVKIKQWSAKYFNGTNI LGRGAGVTLKD 1041
QY 1065 QATNTYFSLVSDNTLPKSLVNPNGHTSSVTGLVFDGKGVYV-YSTSGNQAKNAFISLG 1123
DB 1042 QASNKYLVNDDKLFPLKTLIG-----QVBSGIRPDGTYGTYNSSTTGEKVTDSFITEA 1096
QY 1124 NNWYFDNNGYVMTQAQSGINGANYFLSNGIOLRINAIYDNGNKVLISYDNGRRYENGYY 1183
DB 1097 GNLVYFGDGYVMTQAQSGINGANYFLSNGIOLRINAIYDNGNKVLISYDNGRRYENGYY 1156
QY 1184 LFG-QQWRYFQNGIMAVGLTRVHGVAVQFDASGFOAKGQFIITADGKLYRFDSDGNQIS 1242
DB 1157 QFGNDSWRYFKNGVNALGLTVDGHVQYFDKGVQAKDKIIIVTRDGKVRYPFQHGNGAVT 1216
QY 1243 NRVFVNSKGEWFLPDHNGVAVTGTVTNFORLYFKPNGVQAKGEFIRANGYLRYVYDPNS 1302
DB 1217 NTFVADTKGHWYLYLKGQGVAVTGAOTVKGQHLVFEANGQVKGDFVTAADKGLYFYDVDS 1276
QY 1303 GNEVNRNFRVNSKGEWFLPDHNGVAVTGAOTVKGQHLVFEANGQVKGDFVTAADKGLYFYDVDS 1336
DB 1277 GDMVNTTIEDKAGNWFYLGKDGAAVTAQTQIKG 1310

RESULT 6
GTF2_STRDO
ID GTF2_STRDO STANDARD; PRT; 1592 AA.
AC P27470;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Glucosyltransferase-I precursor (BC 2.4.1.5) (GTF-I) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
OS Streptococcus downei (Streptococcus sobrinus).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OX Streptococcus.
NCBI_TaxID=1317;
RN (1)


```

RESULT 7
GTFD_STRMU
ID GTFD_STRMU STANDARD; PRT; 1462 AA.
AC P49331; O69383; O69386; O69389; O69392; O69398;
DT 01-FEB-1996 (RL 33, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Glucosyltransferase-S precursor (EC 2.4.1.5) (GTF-S) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
GN Names-gtfd; OrderedLocusNames=SMU.910;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=GS-5;
RX MEDLINE=91100958; PubMed=2148600;
RA Honda O., Kato C., Kuramitsu H.K.;
RT "Nucleotide sequence of the Streptococcus mutans gtfd gene encoding
RT the glucosyltransferase-S enzyme.";
RL J. Gen. Microbiol. 136:2099-2105(1990).
RN [2]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=MT4239 / Serotype c, MT4245 / Serotype e, MT4251 / Serotype f,
RC MT4467 / Serotype e, and MT8148 / Serotype c;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
RN [3]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RP STRAIN=UA159 / ATCC 700610 / Serotype c;
RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;
RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,
RA Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
RT pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
CC -!- FUNCTION: Production of extracellular glucans, that are thought to
CC play a key role in the development of the dental plaque because of
CC their ability to adhere to smooth surfaces and mediate the
CC aggregation of bacterial cells and food debris.
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-
CC fructose + (1,6-alpha-D-glucosyl) (n+1).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes
CC water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
CC forms of glucans.
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
CC -!- SIMILARITY: Contains 6 cell wall binding repeats.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL; M29296; AA26895.1; -; Genomic DNA.
CC EMBL; D88653; BAA26103.1; -; Genomic DNA.
CC EMBL; D88656; BAA26107.1; -; Genomic DNA.
CC EMBL; D88659; BAA26111.1; -; Genomic DNA.
CC EMBL; D88662; BAA26115.1; -; Genomic DNA.
CC EMBL; D89979; BAA26121.1; -; Genomic DNA.
CC EMBL; AE014932; AAN58619.1; -; Genomic DNA.
CC HSSP; P06653; 1GYM.
DR InterPro; IPR002479; Cell_wall_bd_put.

```

```

DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 4.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Complete proteome; Dental Caries; Glycosyltransferase; Repeat; Signal;
KW Transferrase.
FT SIGNAL 1 ? Potential.
FT CHAIN ? 1462 Glucosyltransferase-S.
FT REPEAT 1232 1295 1.
FT REPEAT 1296 1359 2.
FT REPEAT 1360 1423 3.
FT REGION 1232 1423 3 X 63 AA approximate tandem repeats.
FT VARIANT 10 10 Y -> H (in strain GS-5, strain MT4239,
FT strain MT4245, strain MT4251, strain
FT MT4467 and strain MT8148).
FT VARIANT 19 19 I -> V (in strain GS-5, strain MT4239,
FT strain MT4245, strain MT4251, strain
FT MT4467 and strain MT8148).
FT VARIANT 58 58 K -> E (in strain MT4467).
FT VARIANT 68 68 A -> S (in strain MT4239 and strain
FT MT4245).
FT VARIANT 81 81 A -> T (in strain MT4251 and strain
FT MT8148).
FT VARIANT 113 113 T -> I (in strain MT4239 and strain
FT MT4245).
FT VARIANT 122 122 A -> V (in strain MT4239, strain MT4245
FT and strain MT8148).
FT VARIANT 132 132 A -> S (in strain GS-5 and strain
FT MT4467).
FT VARIANT 135 135 A -> V (in strain MT4245).
FT VARIANT 137 137 A -> T (in strain GS-5, strain MT4239,
FT strain MT4245, strain MT4251, strain
FT MT4467 and strain MT8148).
FT VARIANT 202 202 V -> L (in strain MT4239).
FT VARIANT 255 255 E -> D (in strain MT4239, strain MT4245
FT and strain MT4251).
FT VARIANT 288 288 D -> N (in strain MT4239, strain MT4245
FT and strain MT4251).
FT VARIANT 301 301 Q -> H (in strain MT4245).
FT VARIANT 313 313 D -> N (in strain MT4239 and strain
FT MT4251).
FT VARIANT 317 317 E -> K (in strain MT4239).
FT VARIANT 328 328 F -> F (in strain MT4239).
FT VARIANT 350 350 F -> L (in strain MT4239, strain MT4251
FT and strain MT4467).
FT VARIANT 628 633 KKKYIQ -> EKEYTL (in strain MT4251).
FT VARIANT 688 688 A -> S (in strain MT4239).
FT VARIANT 726 732 TDQGS -> ADKGNDS (in strain MT4251).
FT VARIANT 726 730 TDQGS -> ADKGN (in strain MT4239 and
FT strain MT4245).
FT VARIANT 762 762 T -> A (in strain GS-5, strain MT4239,
FT strain MT4245, strain MT4251, strain
FT MT4467 and strain MT8148).
FT VARIANT 964 964 D -> Y (in strain MT4251).
FT VARIANT 1019 1019 E -> K (in strain MT4245 and strain
FT MT4251).
FT VARIANT 1059 1060 LG -> IR (in strain MT4251).
FT VARIANT 1060 1060 G -> R (in strain MT4245).
FT VARIANT 1080 1080 G -> R (in strain MT4239).
FT VARIANT 1142 1142 H -> Q (in strain GS-5).
FT VARIANT 1198 1198 S -> N (in strain MT4239).
FT VARIANT 1220 1220 Y -> C (in strain MT4251 and strain
FT MT4467).
FT VARIANT 1280 1280 F -> L (in strain MT4467).
FT VARIANT 1282 1282 Q -> P (in strain MT4245).
FT VARIANT 1290 1290 K -> T (in strain MT4245).
FT VARIANT 1311 1311 G -> D (in strain MT4245).
FT VARIANT 1403 1403 G -> D (in strain GS-5 and strain
FT MT4467).
FT VARIANT 1425 1425 G -> R (in strain GS-5).
FT VARIANT 1449 1449 R -> K (in strain MT4467).
FT CONFLICT 1428 1462 RYDKNSGNMVMKVVTLANGRRIGDRWIARY -> VY
R (in Ref. 1).

```

SQ	SEQUENCE	1462 AA; 163388 MW; CE4A279C4D708645 CRC64;
	Query Match	50.1%; Score 3579.5; DB 1; Length 1462;
	Best Local Similarity	51.9%; Pred. No. 3e-162;
	Matches	711; Conservative 220; Mismatches 375; Indels 63; Gaps 20;
QY	1	MEKKYRFLKRVKRWVTVSIAASAVVTL--TSLGSLVLRKADSTDDRQQAQVTSQASLIVTTS 59
DB	1	METKRYKMYKVKHGWTVTAVASGLITLTGTTLSGSVSAB-----TEQQTSDKVV 51
QY	60	EAAKETLTATDSTATSQPTATVTDNVSTTNGSTNTTANTANFVVKPTTT--SEQA 116
DB	52	QKSEDDKAASE-----SSQTDAPTKQAQTEQQAQSQANVADTSISIKETPSQNIITQA 107
QY	117	KTDSNDKIITTSKAVNRLTATGKFPVANNHTAHPKVTVDKIPIIKPKGLKQPSLSQD 176
DB	108	NSD--DKVTNTKSEBAQSEBRTQAAB--AAQTASSQALTAQAEKLTQKQTAQAEK 163
QY	177	DTAALGNVKNIRKVGKYYKEDGTLOKYNALNTNGKTFPDE--TGALSNNLTSPKGN 235
DB	164	NPVDLAAIPNVKQIDPKYIIGSDGQPKKNFALTNNKLYFPDKNTGALTDSQYQKQG 223
QY	236	IT--NNDNTNSFAQNOVYSTDVANFEHVDHYLTAEISWYRPKYILKDGKTWTQSTEKDPR 293
DB	224	LTFLND-----YTPHQIVNFPENTSELETIDNVTDASWYRPKDIILKNGKXTWTASSEDLR 279
QY	294	PLMTWPDQETQROQVYNNMAQ--LGIHQTYNTATSPLOLNLAQTIQTKIEKITAEN 352
DB	280	PLLSWMPDKQTOIAYLNNMQGLTGENTYADSSQESLNLAQTVQVKIETKISQTOQ 339
QY	353	TNWLQRTISAFVKTQSANSDSEKPD-----DHLQKALLYSNNKLSQANSYRLNR 408
DB	340	TQWLRIINSFVKTPNNNSQTESDSAGEKDHLQGALLYSNSDK--TAYANSRYLLNR 398
QY	409	TPTNQTKDPRYTADRTIGGYEFLANDVNSNPVQAEQLNLHLFMNFCNIYANDPD 468
DB	399	TPTSQTK--PKYFEDNSGGDFLLANDIDNSNPVQAEQLNLHLFMNYSIYANDPE 456
QY	469	ANFDSIRVDADVNDADLLQIAGDYLKAAKGIIHKNDKAANDHLSILEAWSYNDTPYLHDD 528
DB	457	ANFDGVRVADVNVNADLLQIASDYLKAHYGVDSKSEKNAIHLNLSILEAWSDDNPQYKDT 516
QY	529	GDNMNMNRLRLSLYSLAKPLNQ-----RSGWNPILNLSLWRTDDNAETAAPS 580
DB	517	KGAQLPIDNKLRLSLYALTREFLEKDNASKNIRSGLEFVITNSLNNRSAEKGNSERMAN 576
QY	581	YGFIRAHDEVDLRIIRIETINPNVGYSTFTEBEIKAFIYNKDLATKPKYTHYT 640
DB	577	YIFIRAHDESVTVAKIIKAQINPKTDGLTFTLDELKQAFKIYNEDMRQAKKYTQSI 636
QY	641	ALSAYALLTNKSVPRVYGDMDTDDGQYMAHKTINYEAIETLLKARIKYVSGGQAMRNQ 700
DB	637	PTAYALMLNSKDSITRLYYGDMYSDDGQYMATKSPYDAIDTLKARIKYAAGGQDMKIT 696
QY	701	QVGNSE-----IITSVRVGKALKATDGTDRTRTSGVAVIEGNNPSRLKASDRV 751
DB	697	YVEGDKSHMDWDYTGVLTSVRYGTGANEATDQSEATKTQGMVAVITSNPSSLKLNQDKV 756
QY	752	VNNMGAHNKQAVRPLLLTTDNGIKAYHSDQEAAGLVRVTNDRGELIFTAADIKGYANPQ 811
DB	757	IYVMGTAKNQBYRPLLLTLDGLSYTSDAAKSLYRKTNDKGLVFDASDIQGLNPNQ 816
QY	812	VSGYLGVWVPVGAADDQVYRAASTAPSTDGKSVDHQNALDSRVFMFEGSFNPAFKKE 871
DB	817	VSGYLAVWVPVGASDNQDVRVAASNKANATGVYESSSALDSOLIVEGFSNFQDFVKDS 876
QY	872	EYTNVVIANKVDKPAEWGVTDEMAYQYSSYTDGSLDSVIONGYAFTRDYDLIGSKPNK 931
DB	877	DYTNKKIAQNVLQFKSWGVTSEMAPQYVSSDGSFLDSIIQNGYAFEDRYDLAWSKNNK 936
QY	932	YGTADDLVKAIKALHSKGIKMWADWVPDQMYALPKVEVVTATRVDKYGFVAGSOLKNTL 991
DB	937	YGSQDMINAVALKHSGLOVADWVPDQIYNLPKEVVVATRVNDYGEYRDKSEIKNTL 996

QY	992	YVVDKSSGKQQAQYKYGAFLELOAQYPELPARKQISTGVPMDSVKIKQWSAKYFNCT 1051
DB	997	YAANTKSKNGKOYQAKYGGAFLELAAPYSIFNRTQISNGKKIDPSEKITAWEKAKYFNCT 1056
QY	1052	NILGRGAGYVLKDOATNTYFSLVSDNTFLPKSLVNPNGHTSSSVTGLVDFDGKGVVYYS 1111
DB	1057	NILGRGVYVLKDNASDKYFELKGNQTYLPQMTN-----KEASTGFVNDGNGMTFYSTS 1111
QY	1112	GNQAKNAFI--SLGNNNWYFDNNGYVMTGAQSIINGANYFLSNGIQLRNAIYDNGNKVLSY 1170
DB	1112	GYQAKNSFVQDAKGNWYFDNNGHMYGLQHLNGEVQYFLSNGVQLRESFLENADGSKY 1171
QY	1171	YGNDRRYVNGYLLP--CQOWRYFO--NGIMAVGLTRVHGAVOYFPDASGFQAKGQBITTAD 1227
DB	1172	FCHLGNYSYSGYSPDNDKRWYFDASGVMAVGLKTINGNTQYFPDQDGYQVKGAWITGSD 1231
QY	1228	GKLYRFDSDSGNOISNRFVRNSKGWFLFDHNGVAVTGTVTFNGORLYFKPNGVQAKGF 1287
DB	1232	GKRYFDGSGNMAVRFANDKNGDWYLYNSDGLALVGQTINGKTYIFGQDGKQIKGI 1291
QY	1288	IRDANGYLRYDPSNGNEVRNRFVRNSKGWFLFDHNGIAGIATGARVAVNG 1336
DB	1292	ITD--NGLKLYFLANSSELARNIFATDSQNNWYFGSDGVAVTGSQTIAG 1339
RESULT 8		
Q54178	STRGN	
ID	Q54178_STRGN PRELIMINARY; PRT; 1577 AA.	
AC	Q54178_Q54247;	
DT	01-NOV-1996 (TrenBMLrel. 01, Created)	
DT	01-NOV-1996 (TrenBMLrel. 01, Last sequence update)	
DT	01-JUN-2003 (TrenBMLrel. 24, Last annotation update)	
DE	Glucosyltransferase.	
GN	Name=gtfG;	
OS	Streptococcus gordonii.	
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;	
OC	Streptococcus.	
OX	NCBI_TaxID=1302;	
ON	(1)	
RP	NUCLEOTIDE SEQUENCE.	
RC	STRAIN=Challis;	
EX	MEDLINE=96157084; PubMed=8586195;	
RA	Vickerman M.M., Sulavik M.C., Clewell D.B.;	
RT	"Molecular analysis of Streptococcus gordonii glucosyltransferase phase variants.";	
RL	Dev. Biol. Stand. 85:309-314 (1995).	
RP	(2)	
RP	NUCLEOTIDE SEQUENCE.	
RC	STRAIN=Challis;	
RA	Minick P., Vickerman M.;	
RL	Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; U12643; AAC3483.1; -; Genomic_DNA.	
DR	PIR; B41898; B41898.	
DR	HSSP; P06653; 1HCX.	
DR	GO; GO:0016740; F:transferase activity; IEA.	
DR	GO; GO:0009250; P:glucan biosynthesis; IEA.	
DR	InterPro; IPR002479; CW binding.	
DR	InterPro; IPR003318; Glyco_hydro_70.	
DR	Pfam; PF01473; CW binding_1; 5.	
DR	Pfam; PF02324; Glyco_hydro_70; 1.	
KW	Transferase.	
SQ	SEQUENCE 1577 AA; 177805 MW; 5AE0328DC5E08D18 CRC64;	
	Query Match	47.1%; Score 3366.5; DB 2; Length 1577;
	Best Local Similarity	49.6%; Pred. No. 5e-152;
	Matches	695; Conservative 224; Mismatches 398; Indels 85; Gaps 26;
QY	1	MEKKYRFLKRVKRWVTVSIAASAVVTLT----SLGSLVLRKADS----TDDRQQAQVTSQ 52
DB	2	MEKKVHYKMKHKVKNWAIATVTSALLVAPRALGLESQVYADDANQVTNVEQSAVQSK 61
QY	53	ASLVTTSAA-----KEILTATDTSTATSATSQPTATVTDNVSTTNG-----SINTTTA 100

Db 62 DSEQTSDKATDSSQLEVEQASSSKETYQASAAATNPTA-----NEQTQQDKEVETSRD 117
 Qy 101 NTANFVVKPT-TTSEAKTDNSDKIITTSKANRLTA-----TGKRVFVANNNTAHP 150
 Db 118 SRHELTKTSDSSEKSSGSPKQVADQAESTDQTAALQAKQDSRANDQEBETENVA-K 176
 Qy 151 KTVTDKIVPIKPIKIGKLPKS-----SLSQDDIAALGN-----VKNIRKVNKGY 194
 Db 177 ATVSDKII-ATPKERLPEPAQKESITEKMLAAQAQAAPVTEHDDDDVLAKITIDGKK 235
 Qy 195 YYYKEDGTLQKYNALNINGKTFED-ETGAL-SNNTLPSKKNITNNDNTNSFAQVQVY 252
 Db 236 YTVQDDGTGKKNFAVELNGKILYFAETGALVDSNEYQFQCGTSSLN---NEFTQKNAPY 292
 Qy 253 STDVANFEHVDHYLTAESWYRKYILKDGKTWTQSTQKDFRPLMTWPDQETQRYVNY 312
 Db 293 GTTDXDIETVDGYLTADSWYRKYILKDGKTWTQSTQKDFRPLMTWPDQETQRYVNY 352
 Qy 313 MNAQ-LGIHQTYNTATSPQLNLAAQTIQTKIEEKITAEKNTNWLRTISAFVKTSQAWN 371
 Db 353 MNQENLGI-GAPESKTEQVLLTNVAVQVQKIEERISKEGDTKMLRTLMSAFVKTPQNN 411
 Qy 372 -----SDSEKPPDDHLQKGLLYSNNSKLTQANSYRILNRTPTNQTKGKDPRTADRTI 427
 Db 412 IKTESETTGTNKHLOGGGALLYTNSDK-TSHANSRYRLNRTPTSTGTG--TPKYFIDKSN 468
 Qy 428 GGYEFLLANDVNSNPVQAEQLNLWHLPLMNGNIYANDPDANFDSIRVDAVDNVDADLL 487
 Db 469 GGYEFLLANDFNSNPVQAEQLNLWHLPLMNGNISVANDPTANFQGVVDAVDNVDADLL 528
 Qy 488 QIAGDYLKAAKGIIHKNDAANDHLSILEAWSYNDTPYHLDDGDNDNMNDRRLSLLYSL 547
 Db 529 QIADYFYSKYVKGSEBEALKHLISILEAWSNDPDYKDTKGAQLADNKLRLSLLYSF 588
 Qy 548 AKPLNQRGMPLIINSLNVRTDDNAETAAPSYSFIRAHDSVQDLIRNIIRTIINPNV 607
 Db 589 MKLSIRSGVEPTIINSLNDRSTENKNGERTANYIFVRAHDSVQTVIADIIRENINPT 648
 Qy 608 VGYSPTTEIKKAFRIYKOLLATKSKYTHYNTALSYALLTNKSSVRVYVYVGMFTDGG 667
 Db 649 DGLTITMELQKAFKINEDMRKADKKYQFNIPTAHALMLSNKDSITRVYVGDLYTDDG 708
 Qy 668 QYMAHTINYEAIETLLKARIKYVSGGQAMRQVQ-----NSEIITSVRVYVYVGMFTDGG 718
 Db 709 QYMEKKSYPHDAIDALLRARIKYVAGGQDMKYTVNGVPREADKWSYNGILTSVRVYVGMFTDGG 768
 Qy 719 KATDTGDRTRTSQVAVIEGNNPSURLKASDRVVVNMGAHKKQAYRPLLLTDTNGIKAY 778
 Db 769 EATDGTGTAETRTQGMVAVIASNNPNLKLNEWDKLVNMGAAHKNQYRPLVLLTKDGIKRY 828
 Qy 779 HSDQEA-AGLVRYTNDRGELIFTAADIKGYANPOVSGYLVVPGVGAADQDVRVAASTA 837
 Db 829 LTDEVPQSLMKTDTANGILTFDMNDIAGYSNVQVSGYLVVPGVGAADQDVRVAASTA 888
 Qy 838 PSTDCKSVHQAALDSRVWFEGSFNFQAFATKKEBYTNVJAKGVYDKEAEGVTFDEMAP 897
 Db 889 KNASQVYVESPALDSQLIYEGSFNFQAFATKKEBYTNVJAKGVYDKEAEGVTFDEMAP 948
 Qy 898 QYVSTGDSFLDSVITQNGYAFDTRDYLGIKSPNKYGTADDLVKAIKALHSGIKVYMAOW 957
 Db 949 QYVSSQDGTFLDSIIQNGYAFEDRYDMAMSKNKKYGLSLDILLNLRALHLSVNIQAIADW 1008
 Qy 958 PQMVALPEKEVWATRVYDGTGTPVAGSQIKNTLYVVDGKSGQDQQAQYGGAFLEBLOA 1017
 Db 1009 PDQIYNLPKGVWATRVYDGTGTPVAGSQIKNTLYVVDGKSGQDQQAQYGGAFLEBLOA 1068
 Qy 1018 KYPELFARKQISTGVPMDDPSVKIKOWSAKYFNGTNILGRGAGYLVKDAQATNTYPSLVSDN 1077
 Db 1069 KYPELFARKQISTGVPMDDPSVKIKOWSAKYFNGTNILGRGAGYLVKDAQATNTYPSLVSDN 1128
 Qy 1078 TFLPKSLVNPNGHTSSSYTGLVDFGKGYVYVYSTSGNQAKNAFISLGN-NWYVFDNNGYV 1136

Db 1129 TALPQKLVN-----KEASTGFVKDTNGFKFYSTSGNQAKDTFIQDENGNNWYFDNQGLV 1183
 Qy 1137 TCAQOSINGANYVFLNGIOLRNAIYDNGKNKLSYTGNDGRRYRVEGYLFL-GQOMRYFQ-N 1194
 Db 1184 TGAREIDGKQLYFMKNGVQLRDALQEDENGQYVYDQTKGAKVLNRYTSDGQNWRYFDAK 1243
 Qy 1195 GIMAVGLTRVHGAOVYFDASGFOAKGQFITTADGKLYRFDSDSGNQISNRFRV-NKSGEW 1253
 Db 1244 GYMARGLVKIGDQYVFDQNGYQVKGKVRADKGLRYFDKDSGNVAVINRFAQGDNPDSW 1303
 Qy 1254 FLFDHNGVAVTGTVTTFNGQRLYFKNGVQAKGEFTDANGLYRYDYPNSGNEVNRNFRV 1313
 Db 1304 YTFGADGVALTQLQKIGQQTLYFGQDGKQVQVVLADKRSIRYFDANSGEAVNFAEG 1363
 Qy 1314 SKGEWFLPDHNGIAVTGARVNN 1335
 Db 1364 AKNEWYFFDQDGKAVTGLKTN 1385

RESULT 9
 Q69A94 LEUME
 ID Q69A94 LEUME PRELIMINARY; PRT: 1454 AA.
 AC Q69A94;
 DT 25-OCT-2004 (TremBLrel. 28, Created)
 DT 25-OCT-2004 (TremBLrel. 28, Last sequence update)
 DE Dextranucrase (EC 2.4.1.5).
 GN Name=darP;
 OS Leuconostoc mesenteroides.
 OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
 OX NCBI_TaxID=1245;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=IBT-PQ;
 RA Fernandez-Vazquez J.L., Lopez-Munguia A., Olivera C.;
 RT "Molecular characterization of a dextranucrase gene from Leuconostoc
 mesenteroides IBT-PQ isolated from pulque."
 RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY504865; AAS79426.1; -; Genomic DNA.
 DR GO; GO:0047849; F:dextranucrase activity; IEA.
 DR GO; GO:0004675; F:transferase activity, transferring glycosyl. . .; IEA.
 DR GO; GO:0009250; P:glucan biosynthesis; IEA.
 DR InterPro; IPR002479; CW binding.
 DR InterPro; IPR003318; Glyco_hydro_70.
 DR Pfam; PF01473; CW binding_1.
 DR Pfam; PF02324; Glyco_hydro_70; 1.
 KW Glycosyltransferase; Transferase.
 SQ SEQUENCE 1454 AA; 160569 MW; 2DC7342963B6FD87 CRC64;

Query Match 46.9%; Score 3348.5; DB 2; Length 1454;
 Best Local Similarity 51.6%; Pred. No. 3.2e-151;
 Matches 699; Conservative 195; Mismatches 392; Indels 69; Gaps 27;

Qy 17 VTVSIASAVVLTLSLGSGLVKADSTDDRRQAVTESQASLVTTSEAAKETLTATDSTATS 76
 Db 32 VTSPIQQAQADVQSKNGVVVIT-----AVNQSNDAITTD-----KSIITNDKATTTA 79
 Qy 77 ATS-OPATVTDNVSTTQNTTANTANFVVKPTTSEQAKTDSKLIITTSKAVNRLT 135
 Db 80 DTSTNDKATTTADTSTNDKATTTADTSTN--DKATTTAD--ISTNNKATTTADTSTNNKA 135
 Qy 136 ATGCFVPPANNNTAHPKTVTDKIVPIKPIKIGKLPQSPSSISQDDIAAL-GNVKNIRKVNKGY 194
 Db 136 ATTADTSDNNNSA---TTSDKDVSSLAOKSAIDNNKSTTTDAAFEASSKMLKTDGKT 192
 Qy 195 YYYKEDGTLQKYNALNINGKTFED-ETGALSN-NLPSKKNITNNDNTNSFAQVQVY 252
 Db 193 YYYDNGQIKKEATVIDGKLYFDKDTGALADTNDYQFLEGLTSEN---NNYTEHNASV 249
 Qy 253 STDVANFEHVDHYLTAESWYRKYILKDGKTWTQSTQKDFRPLMTWPDQETQRYVNY 312
 Db 250 GTSSASVNTVDGYLTADSWYRKYILKDGKTWTQSTQKDFRPLMTWPDQETQRYVNY 306

QY 479 VDNVADLLOIAGDYLKAAKGIHKNDKAAANDHLSILEAWSYNDTPYLHDDGDNNMNDNR 538
Db 518 VDNVADLLOIAGDYLKAAKGIHKNDKAAANDHLSILEAWSDNDPDYNDKOTKGAQLAIDNK 577
QY 539 LRLSLLYSLAKPLNQSGMPLTNSLVNRTDNDNAETAAVPSYFTRAHDSEVQDLIRNI 598
Db 578 LRLSLLYSPWNLISIRSGVEPTITNSLNDSSSEKNGERMANIFVRAHDSEVQTVIADI 637
QY 599 IRTEINPNVGVSTTEBIKAFIYNNKOLLATEKKYTHYNTALSALLTNKSSVPRVY 658
Db 638 IRENINPNVGTDTMDELQAKFIYNEDMRKADKKYQFNIPTAHALMSLNKSDITRVY 697
QY 659 YGDMFTDDGOYMAHKTIYEAETILKARIKYSGGAMRNOQVG-----NSEIIT 709
Db 698 YGDLTYDDGGYMKKSPYDAIDALLARIKYVAGQDMKYTMGVFPREADKWSYNGILT 757
QY 710 SVRYGKALKATDGTORTTTSVAVTEGNNPRLKASDRVVVNNGAHKNQAYRPLL 769
Db 758 SVRYGTGANEATDEGTAETFTQMAVIASNNPRLKNEWDKLVNNGAHHKNQYRPLL 817
QY 770 TTONGIKAYHSDQEA-AGLVRYNDRGELIFTAADIKGYANPOVSGYLVVYVPGAAADQ 828
Db 818 TTQDGISRYLTDEVPQSLWKKTDANGILTFDMNDIAGYSNVQVSGYLAWVYVPGAKADQ 877
QY 829 DTVRAASTAPSTDCKSVHQAALDSRVMPGEPFNFQAPATKKEEYTNVVIKNDVKAEM 888
Db 878 DARTTAKSKKNASGVYESSAALDSQLIYEGFNFQDFATRDQDQYTNKVIAKNVLFKEW 937
QY 889 GVTDPMAPOYSSYSDGSLDSVIQNGYAFTRDYDIGISPKNKYGTADDLVAKALHSK 948
Db 938 GVTSELPPOYSSQDGTFLDSIIQNGYAFEDYDWMKNNKNGISLKKOLLNLRALHSV 997
QY 949 GIKVMADWPQOMYALPEKVEVVTATRVKYPVAGSQIKNTLYVVDGSSGKQDQAKY 1008
Db 998 NIQAIADWPQOYINLPKGVVVTATRVNNGYVREGAEIKEKLYVANSKTNETDFGKY 1057
QY 1009 GAFLPELOAKYFELPARKQISTGVPMDSVKIKQWSAKYFNGTNIILGRGAGYVLKQDQATN 1068
Db 1058 GAFLDELKAKYFELPARKQISTGVPMDSVKIKQWSAKYFNGTNIILGRGAGYVLKQDQATN 1117
QY 1069 TVFSLVSDNTELPKSLVNPNGHTSSSVTGLVDPGKGYVYVSTGNOAKNAPISLGN-NKY 1127
Db 1118 DYLTVRNGEIVLPKLVNKN-----SYTGVSANGTKFYSTGYQAKNSFIDENGWY 1172
QY 1128 YFDNNGYMYTGAQSGINGANYFSLNGIQLRNAIYDNGNKVLSYVYNGDGRYENGYYLF-G 1186
Db 1173 YFDKRGYLVGAHEIDGKHVFLKNGIQLRDSIREDENGQYVYDQGAQVLRYYTTDG 1232
QY 1187 QOWRYFO-NGIMAVGLTRVHGAQVYFDSAGPOAKQFITTADGKURYFDRDSGNQISNRF 1245
Db 1233 QNWRYFDAGVNMARGLVKIGDQGFDENGYQVKGKIVSAKDGKLYRFPKDSGNVIRNF 1292
QY 1246 VR-NSKGWFLPDHNGVATGTVTFNGQRLYKPKNGVQAKGEFIRDANGLYRYYPNSNG 1304
Db 1293 AGDNPSDMYFVGVFAKLTGLQKTQQOYLYFDQDGKQVKGKIVLSDSKSIRYFDSNGE 1352
QY 1305 EVRNERFVNSKGEWFLPDHNGIATGARVV 1334
Db 1353 MAVGFAEAGKAEWYFDTGKAVTGLQKI 1382

RESULT 11

Q55265 STRSL
ID Q55265 STRSL PRELIMINARY; PRT; 1577 AA.
AC Q55265,
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Glucosyltransferase precursor.
GN Name=gtfM;
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.

OX NCBI_TaxID=1304;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=95122197; PubMed=7822030;
RA Simpson C.L., Giffard P.M., Jacques N.A.;
RT "Streptococcus salivarius ATCC 25975 possesses at least two genes
RT coding for primer-independent glucosyltransferases.";
RL Infect. Immun. 63:609-621(1995).
DR EMBL; L35928; AAC41413.1; -; Genomic_DNA.
DR PIR; T30858; T30858.
DR HSSP; P06653; 1H8G.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR004829; C:surface antigen.
DR InterPro; IPR002479; C:surface antigen.
DR InterPro; IPR003318; G:glycohydro_70.
DR Pfam; PF01473; C:binding_1; 4.
DR Pfam; PF02324; Glycohydro_70; 1.
DR ProDom; PD153432; C:surface antigen; 1.
KW Signal; Transferase.
FT SIGNAL 1 38 Potential.
FT CHAIN 39 1577 Glucosyltransferase.
SQ SEQUENCE 1577 AA; 175290 MW; 3EFB898A7D3A7BF3 CRC64;

Query Match 45.0%; Score 3216; DB 2; Length 1577;

Best Local Similarity 47.7%; Pred. No. 7.7e-145;

Matches 687; Conservative 214; Mismatches 413; Indels 126; Gaps 33;

QY 1 MEKVRPKLRKVRKRWTVSIAAVVTLTSLG-----SLVKADST-----DDRQAVT 49
Db 1 MENKVRPKLRKVRKRWTVSIAAVVTLTSLG-----SLVKADST-----DDRQAVT 58

QY 50 ESOASLVTS-----EAAKETLTATDTST-----ATSATSO-----PAT-- 84
Db 59 DGTASLVTS-----EAAKETLTATDTST-----ATSATSO-----PAT-- 118

QY 85 ---VTDNVSTNTSTNTT-----ANTANFVVKPT-----TTSQAK-----TDNSDKIITS 128
Db 119 VAVSSQTSSQSGQETTEQVSGQSTQVAGTSAQSTPSVTQARPRVLTNAAPALATR 178

QY 129 KA-----VNRLTATGKFPVANNHTAHKTVTKIVIPKPKI----- 164
Db 179 AADSTIRINANNNTNITITASGTTTNPVTIITGPNT- KPNVTVTSNTPNVTIITQPN 237

QY 165 --GKLKPSLSQDDIAALGN-----VNIRKNGKYYVYKEDGLTKNYALNNG 213
Db 238 QNKPVPQSPQSPKPKVPQNPQPSLDYKPVASNLKTIDGKQYV-ENGVVKNAALELDG 296

QY 214 KTFEFDGALSNNTLPSKKGN-ITNNDNTNSFAQYNQVYSTDVANFEHVDHYLTAESWY 272
Db 297 RLYVFEDETGAMVDQSKPLVRAAI PNN---SIYAVYNOAYDTSSKSEFHLDNFLTADSWY 353

QY 273 RPKYILKDGKWTQSTEDKDFRPLMTWPDQSTQRYVNMNAQLGIHOIYNTATSPQL 332
Db 354 RPKQLKDGKWTQSTEDKDFRPLMTWPDQSTQRYVNMNAQLGIHOIYNTATSPQL 413

QY 333 NLAQOTIQTKTEEKITAENKTNMLROTISAFVKTSQANSDSEKPP---DDHLOKALLY 389
Db 414 AAAAETVGRGEERIGREGNTTLWLRLQMSDFITQPGMNSSEEDNLLVKGKHLQGGALTF 473

QY 390 SNNSKLTQANSNYRILNRTPTNQTGKPKDPRYATDRTIGGYEFLFLANDVNSNPVQAEQ 449
Db 474 LNNSS-ATSHANSDFELMNRTPNTQGTGR--KXIDRSNGGYELLANDIDNSNPVQAEQ 530

QY 450 LNLWHLFANFGNIYANDPDANFDSIRVDVNDVADLLOIAGDYLKAAKGIHKNDKAND 509
Db 531 LNLWHLFANFGNIYANDPDANFDSIRVDVNDVADLLOIAGDYLKAAKGIHKNDKAND 590

QY 510 HLSILEAWSYNDTPYLHDDGDNNMNDKRLSLILYSLAKPLNQSGMPLTNSLVNRT 569
Db 591 HLSILEAWSYNDTPYLHDDGDNNMNDKRLSLILYSLAKPLNQSGMPLTNSLVNRT 650

QY 570 DDNAETAAPSVSYFTRAHDSEVQDLIRNIIRTEINPNVGVSTTEBIKAFIYNNKOLL 629


```
Db      866  VQDAQYTNKKIAENTDLFKAWGVTSFEMAPQVVSATDGTFLDSIIQNGYAFSDRYDLAM 925
Qy      927  SKNKYGTADDLKAIALKHSKGIKVMADWVPDQYALPEKEVVTATRVDKYGTVPAGSQ 986
Db      926  SKNKYGSKEDLANALKAHAGIAIADWVPDQIYQLPEKEVVTASRVNDYGRVKIDQP 985
Qy      987  IKNTLVVDPGSKGDKQQAQYGGAFLEELQAKYPELPARKQISTGVPMDPSPVKIKOWSAK 1046
Db      986  MVNKLVLANTKSSGKDFQAKYGGEFLEELQAKYPELPARKQISTGVPMDPSPVKIKOWSAK 1045
Qy      1047  YFNGTILGRGAGYVLKQATNTYFSLVSDNTFLPKSLVNPNGHGTSSSVTGLVDFGKGVY 1106
Db      1046  YFNGTILGRGAGYVLKQATNTYFSLVSDNTFLPKSLVNPNGHGTSSSVTGLVDFGKGVY 1106
Qy      1107  YFNGTILGRGAGYVLKQATNTYFSLVSDNTFLPKSLVNPNGHGTSSSVTGLVDFGKGVY 1107
Db      1107  YFNGTILGRGAGYVLKQATNTYFSLVSDNTFLPKSLVNPNGHGTSSSVTGLVDFGKGVY 1107
Qy      1102  YFTAGSQAQKSFVTVAGTYFYDYTGHWVTGNGINTFYFPLPGLVGLVTHGNVQYIDEE 1161
Db      1102  YFTAGSQAQKSFVTVAGTYFYDYTGHWVTGNGINTFYFPLPGLVGLVTHGNVQYIDEE 1161
Qy      1165  NKVLSYNGDGRYE----NGYLF----GQ-QMRYFON-GIMAVGLTRVHGAVQYFDAS 1214
Db      1162  RSV--YYGKTGVNKGSRNWFAMTDSKGLRFRHFDNYGFMVGLVTHGNVQYIDEE 1219
Qy      1215  GFQAKQFITTADGKLRVFDKSGNQISNRFRVNSKGEWFLPDHNGVAVTGVTNGQRL 1274
Db      1220  GFQAKQFITTADGKLRVFDKSGNQISNRFRVNSKGEWFLPDHNGVAVTGVTNGQRL 1274
Qy      1275  YF-KPENGVAQGEFIRANGLYRYDPNSGNEVRNFRVNSKGEWFLPDHNGVAVTGARV 1333
Db      1279  YFADKTAQVKGDFVTDKDGNTFFYSGDGTGLAVSTFFSTGNAMPYADENGHVAKGEKT 1338
Qy      1334  VNG 1336
Db      1339  ING 1341

RESULT 13
ID Q56CX8_9STRE PRELIMINARY; PRT; 1506 AA.
AC Q56CX8;
DT 10-MAY-2005 (TremBLrel. 30, Created)
DT 10-MAY-2005 (TremBLrel. 30, Last sequence update)
DE 10-MAY-2005 (TremBLrel. 30, Last annotation update)
DE Glucosyltransferase-T.
GN Name=gtft;
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1310;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=B13N;
RA Kuwahara N.;
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY966490; AAX76986.1; -, Genomic_DNA.
KW Transferase.
SQ SEQUENCE 1506 AA; 167016 MW; 94B882EF2C17C451 CRC64;

Query Match 44.5%; Score 3175.5; DB 2; Length 1506;
Best Local Similarity 46.6%; Pred. No. 6.2e-143;
Matches 645; Conservative 237; Mismatches 404; Indels 99; Gaps 25;

Qy      1 MEKKVRPKLRKVRKRWTVSTASAVVLTLSGSLVKADSTDD--RQAVTESQASLVTT 58
Db      1 MERKLYKLRKVRKRWTVSTASAVVLTLSGSLVKADSTDD--RQAVTESQASLVTT 58
Qy      59 SE-----AAKETLTATDTSATSSQPTATVTDNVSTNTSTNTTANTFVVK----- 108
Db      61 NEDEVASDAADTASAKATSEKEVQSSDTSNSETNQVETKQQA--SAKESADAVAKAPQA 118
Qy      109 -PTTTSQAKTDSNKIITTSKAVNRLTATGKFPANNNTAHPKTVTDKIVPKPKIGKL 167
Db      119 GPATTSQVASSSESS--VAPSKADKAAA----- 145
```

```
Qy      168  KOPSSLSODDIAALGNVKNIRKNGKYVYKEDGTLLQKNVALNINGKTFEED--ETGALS 226
Db      146  ---GSVQNEEERAAISLANIKKIDGYITVMDAGSYKCNFAITVDCQMLTFDAKTALSS 202
Qy      227  NTLPSKKGITNDNTNTPAQYQVYSTDVANFEHVDHYLTAEBSWYRPKYILKDGKTWQ 286
Db      203  TSYTSPSQGLT--PIVSDFSVNNKAFDSSEKSPFELVDGYLTAEBSWYRPKILENGKTW 260
Qy      287  STEKDFRPLMTWPDQETORQYVYNNMAQLGHQHTYNTATSPQLNLAAQTOTKIEEK 346
Db      261  SKETDLRPVLMGWPNKQTVAYLNTMSKALGKEEFTTETSQLTTLTAELIAQKIEAR 320
Qy      347  ITAEKNTNLRQITSAFVKTSQAMNSDSEKPPD--DHLQKGLLYNNNSKLTQSAANSYR 404
Db      321  VSEKQGTWLRAMAFVATQSRWKKDSEQ--YDKADHLOGGALLYTNNN--LTEWANSNR 378
Qy      405  ILNRTPTNQTGKDPRTYATDRITIGGYEFLLANDVNSNFPVQAEQLNLHFLMNFNIA 464
Db      379  LLNRTPTRQDQK--THYSKADKYGGYEFLLANDVNSNFPVQAEQLNLHFLMNFNIA 436
Qy      465  NDPDANFDSIRVDVNDVADLLOIAGDYLLKAAKGIHKNDKAAHDLHLSILEANSYND 524
Db      437  GDNANFDSIRVDVNDVADLLOIAGDYLLKAAKGIHKNDKAAHDLHLSILEANSYND 496
Qy      525  LHDDGDNMNMNRRLSLLSYLAQPLNORS--GMPLITN--SLVNRITDDNAETAAPSY 581
Db      497  NQDTNGAALAMNDGRLFSLLYTLTPINERTFGMSTLIKSEYGLDTRTKNDKYGDTQPS 556
Qy      582  SFIRAHSEVDLIRNIIRTEINPNVGVSYFTTBEIKKAFIYNKDLATEKRYTHNTA 641
Db      557  VEVRAHSEVDQVIAQIIKEIDPTDGTFTLDQLKQAFIYNKDMNSVKNHYTHNIP 616
Qy      642  LSYALLLNKSSVPRVYGDMTDDGQYMAHTINYAEIETLLKARIKTVSGQAMNQ 701
Db      617  AAYAVMLSNMESVTRVYGDFTDQGYMAKSPYDAINTLLRARIRYAAGQIMEHNS 676
Qy      702  -----VGNSEITSVRYGKALAKATO--TGDRTTTSVAVTEGNNPSL 743
Db      677  YKPSAAMKAAPHDAGNVLGNSVLVSVRFGQVMSADMTGKGLAKTSQMTLISNNPEL 736
Qy      744  RLKASDRVVVNMGAHKNQAVRPLLLTTDNGIKAYHSDQEAAGLVRYTNDRGELIPTA 803
Db      737  ELDVNEEIKVNVGKHAGQAVRPLLLTTDNGIKAYHSDQEAAGLVRYTNDRGELIPTA 795
Qy      804  IKGYANPOVSGYLVGVVPGAAADODVRAASTAPSTD--GKSVHQAALDSRVMEGFG 862
Db      796  IKGYKQVEVNGYLSVWVPVGAKDQDIRVAPSTAARKEKAKTYTASQALLESQLIYEG 855
Qy      863  FOAPATKKEEYTNVVIKNDKFAEWGVTDPFMAPOYVSSTDGSFLDSVIONGYAFTDR 922
Db      856  FQDFVQKDSQYTNKKIAENTDLFKAWGVTSFEMAPQVVSATDGTFLDSIIQNGYAF 915
Qy      923  DLGSKPNKYGTADDLKAIALKHSKGIKVMADWVPDQYALPEKEVVTATRVDKYGT 982
Db      916  DLAMSKNNYKSKEDLANALKAHAGIAIADWVPDQIYQLPEKEVVTASRVNDYGRVK 975
Qy      983  AGSQIKNTLYVVDGKSSGKDKQQAQYGGAFLEELQAKYPELPARKQISTGVPMDPSP 1042
Db      976  VDQPLVEKLYLANTKSSGKDFQAKYGGEFLEELQAKYPELPARKQISTGVPMDPSP 1035
Qy      1043  WSQKYNFNTNLTGRGAGYVLKQATNTYFSLVSDNTFLPKSLVNPNGHGTSSSVTGL 1102
Db      1036  WSQKYNFNTNLTGRGAGYVLKQATNTYFSLVSDNTFLPKSLVNPNGHGTSSSVTGL 1091
Qy      1103  KGYVYVYSGNQAKNAFISLGNWYFIDNNGYVMTVTAQSGINGANYFSLNGIQLRNAI 1162
Db      1092  KGIYVYVYTAGNKARSFAFTEAGTYFYDYTGHWVTGNGINTFYFPLPGLVGLVTHGN 1151
Qy      1163  NGNKVLSYNGDGRYE----NGYLF----GQ-QMRYFON-GIMAVGLTRVHGAVQY 1212
Db      1152  DEKRSVYVYKGTGVNKGSRNWFAMTDSKGLRFRHFDNYGFMVGLVTHGNVQYIDEE 1211
```



```
QY 1213 ASGFOAKGOFITADGKLYRDRDSQNIISNRFVRSKGWFLPDHNGVAVTGTTFNGQ 1272
Db 1212 ENGFOVKGEFVDQDQTRYPDQGGNLVKGFL-NKDGWYLYDDQGLVAKGAQTIGQ 1270
QY 1273 RLYF-KPNGVQAKGFIIRDANGYLRYDPSNGEVRNRFVRNSKGWFLPDHNGIAVTGA 1331
Db 1271 KLYFDTKGVQVKGDFVTDKNGTTFYSGDTGLILGQFFSTGNNAFVADENGHVAKGA 1330
QY 1332 RVNG 1336
Db 1331 KTRIG 1335

RESULT 14
ID Q8KZL5_9STRE PRELIMINARY; PRT; 1554 AA.
AC Q8KZL5;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Glucosyltransferase.
GN Name=gtfU;
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1310;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21958684; PubMed=11960691; DOI=10.1016/S0304-4165(01)00240-9;
RA Hanada N., Fukushima K., Nomura Y., Senpuku H., Hayakawa M.,
RA Mukasa H., Shiroza T., Abiko Y.;
RT "Cloning and nucleotide sequence analysis of the Streptococcus
RT sobrinus gtfU gene that produces a highly branched water-soluble
RT glucan."
RL Biochim. Biophys. Acta 1570:75-79(2002).
DR EMBL; AB089436; BAC07265.1; -; Genomic_DNA.
DR HSP; P06653; 1H8G.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW binding_1; 5_70.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1554 AA; 171676 MW; 6981BCC1DAE24A73 CRC64;

Query Match 44.3%; Score 3161.5; DB 2; Length 1554;
Best Local Similarity 48.5%; Pred. No. 3e-142;
Matches 662; Conservative 223; Mismatches 400; Indels 81; Gaps 31;

QY 1 MEKKVRFKLKRVKKRWTVSIVASAVVLTLSGSLVKADSTDDROQAVTESQASL----- 55
Db 1 MEKLLHYLKHVKKGHWVTIAVAS--IGLSLVGA--GTVSABDKVANDTTAQATVGVDTG 56
QY 56 ---VTTSEAKETLTATDTSTATSOPTATVTNVSITNSTNTANTANFVVKPTT 112
Db 57 QDQATNDANTNT-TDITDQASANTNQDQAG-SQSNQDQAKQDQANT-----DRNQ 108
QY 113 SEQAKTDSNCKIITTSKAVNRLTATGKFPANNNTAHPTVTDKIVPKPKIGKLGKQPSS 172
Db 109 ADNSQTDN-----NQATDQATS-----PATDGTISVQRDAANVATAADQEG---QTAP 153
QY 173 LSQDDIALGNVKNIRKNGKYKYYKEDGTLOKNVALNINGKTFPFD-ETGALSNTPLPS 231
Db 154 SEQEKSAAAL-SLDNVKLIDGKYTVVQADGSYKKNFAITVNGQMLYFSDTSGALSTSTYS 212
QY 232 KKGNTINDNTNSFAQYNQVGTVDVANFEHVHLYLTAEWSYRPXYLTKDGKTWTOSTEKD 291
Db 213 FSQGTIN--LVDDFSHKNKAYDSTAKSPFLVNGYLPANWSYRPAGLRNGQTWEASNND 270
QY 292 FRPLMTWMPDQETQRQYVYNNQAQIGHOTYNT-ATSPLOLNLAQAQTIQKIEKITAE 350
Db 271 LRFVLMSWPFDKDTQVAVYVNNKYLSANETETVNETSQVDLNLKEAQS!QTKIEKITS 330
```

RESULT 15
Q55264_STRSL
ID Q55264_STRSL PRELIMINARY; PRT; 1449 AA.

```
QY 351 KNTWNLROTIISAFYKTSQSAWNSDSEKFPD--DHLOKGALLYSNNSKLTQANSNYRIINR 408
Db 331 NSTOWLRITAMEAFVAAQPKWNNSTEN-FNKGDHJUGGALLYT-NSDLTPWANSRLLNLR 388
QY 409 TPTNQTGKDDPRYTADRTIGGYEFLLANDVNSNPVQAEQLNMLHFLMNFNFIYANDPD 468
Db 389 TPTQODGTK--KYFTEGEGGYEFLSNDVNSNPVQAEQLNQLHLYLNMWGDIVMGDKD 446
QY 469 ANFDSIRVDVNDVADLLOJAGDYLKAAGIHKNDKKAANDHLSILEAWSYNDTPLYHDD 528
Db 447 ANFDGVRVDVNDVNADLLQVYSNYFKDNYKVTDSANALAHISILEAWSLNDNQYNET 506
QY 529 GDNMLNDRRLRLSLLSYLAFLNORSQWNPILITNSL-VNRDQNAETAAPVSYSFIRAH 587
Db 507 NGTALSIDNSRLTLAVLTQPGQRIDLSNISVESVKNERANDTAYGDTITFYFVRAR 566
QY 588 DSEVQDLIRNIIRTEINPNVVGYSPTTEEIKKAFIYNKDLIATEKKTYHTNTALSYALL 647
Db 567 DSEVQTVIAKIVKEKIDTNSDGYTFLDLQKDAFKLYNEDMAKVNKTYTHYNIIPAAYALL 626
QY 648 LTNKSSVPRVYVYGDFTDDGQYMAHKTINYBAIETLLKARIKYVGGQAMRNQOV-GNSE 706
Db 627 LSNMESVPRVYVYDLYTDDGQYMAKKSPPYDAIATMLQGRYAYVSGGQSEEVHKVGNNG 686
QY 707 IITSVRYGKGALKATDT-GDRTTTSQVAVIEGNNPSLRLLKASDRVVNMGAHKNQAYR 765
Db 687 ILSSVRYGQDLMSADDDTQGLDLSRTSGVLTIVSNDPNIDL-GGDSLTVMNGRAHANQAYR 745
QY 766 PLLLTNDNGIKAYHSDQEAAGLVRYTNRDGLIFTAADIKGYANQVSGYGLGVWVPVGA 825
Db 746 PLLLTGKGVQSYLKDS-DTHVKTVDANGNLTFADDIKGYSTVDMSGYLAVWVPVGA 804
QY 826 ADQDVRVAASTAPSTDGKSVHQNAAALDSRVMPGFSNFQAFKTKKEEYTNVVIKNVDKF 885
Db 805 DQDQVRVAADTNQKADGKSLKTSAAALDSQVIEGFSNFQDFANNDADYTNKKIAENADFF 864
QY 886 AEWGVTDFEMAPQVYSSDTGSLDSVIONGYAFTDRYDLGISTKPNKYGTADDLVAKAL 945
Db 865 KKLGITSFEMAPQVYVSATDGSFLDIIQNGYAFSDRYDLAMSKNNKYGSKDDLALAKAL 924
QY 946 HSKGIKVMADWVPDQMYALPKEVVTATRVKGYTPVAGSQIKNTLYVVDGKSSGKQQA 1005
Db 925 HANGIOAITADWVPDQIYQLPGEVVTAKRYSYGNPTFDAYINNALYATNTKSSGSDYQA 984
QY 1006 KYGGAFLBEELOAKYPELPFARKQISTGVPMDSVKIKQWSAKYFNGTNIILRGAGYVYKQ 1065
Db 985 QYGGAFLDLAKAKYPMFTVNMISTGKPIDPSTKIKQWEAKYFNGTNIILRGAGYVYKQ 1044
QY 1066 ATNTYFSLVSDNTLPSLVNPNHGTSSSVTLGVPDGGYVYVYSTSGNOAKPAFISLGN 1125
Db 1045 ATGYFTYVNEGDFLPASFT---GDQNAKTGYFDGTGMAYYSTSGNAKVNFSIYEGGH 1100
QY 1126 WYFONNGYMTGA-QSINGANYFPLSNGIQLRNAIYDNGNKNVLSYGYNDGRRYENGYYL 1184
Db 1101 YYYFDKGHWYTGSKAEDGNDYFELPNGIQNRDAIYQDAQGSYYYGRTGLYK----- 1155
QY 1185 FQQW-----RYFQ-NGIMAVGLTRVHGAVQVYPDASGFAKQGFITADOK 1229
Db 1156 -GDNYFPVDPNNANKTVFRYFDANNVMAIGYRNMYGQTYYPDENGFQAKGQLLTDKQ- 1213
QY 1230 LEYFDRDSGNQISNRFVRSKGWFLPDHNGVAVTGTTFNGORLYEKP-NGVOAKGETI 1288
Db 1214 THYFEDNGAMAKNKFV-NVGDDWYMDGNGNAVKGQVFNQIILYFNDETGVQVKGQFI 1272
QY 1289 RDANGLYRYDPSNGEVRNRFVRNSKGWFLPDHNGIAVTGARVV 1334
Db 1273 TDAQRTSYDANSALKSSGFFTPNGSDWY-YAENGYYVYKQKV 1317
```

```
AC Q55264;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Glucosyltransferase precursor.
GN Names:gtfII;
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=95122197; PubMed=7822030;
RA Simpson C.L., Giffard P.M., Jacques N.A.;
RT "Streptococcus salivarius ATCC 25975 possesses at least two genes
RT coding for primer-independent glucosyltransferases.";
RL Infect. Immun. 163:609-621(1995)
DR EMBL: L35495; AAC41412.1; -; Genomic_DNA.
DR FIR; T30857;
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
DR Signal; Transferase.
KW SIGNAL
FT CHAIN 1 35 Potential.
FT CHAIN 36 1449 Glucosyltransferase.
FT CHAIN 1449 AN; 159984 MW; DD62F07306E86A46 CRC64;
SQ SEQUENCE 1449 AN; 159984 MW; DD62F07306E86A46 CRC64;

Query Match 44.2%; Score 3159; DB 2; Length 1449;
Best Local Similarity 47.0%; Pred. No. 3.6e-142;
Matches 669; Conservative 197; Mismatches 439; Indels 118; Gaps 28;

QY 1 MEKKYRFKLRKVKKRWTVSI-----ASAVTLTSLSLVKADSTDRQAAVTESSQAS 54
DB 1 MEKKYRFKLRKVKKRWTVSI-----ASAVTLTSLSLVKADSTDRQAAVTESSQAS 54
QY 55 LVTTSEAAKETLTATDTSTATSQS---PTATVDNVSTTQSTNTTANTANFVVKPTT 111
DB 57 TVDTGTVSDNTTAAQDPTTAAVATNDVATDQATPTATFLLTDTTNTVAANAVDVATVG 116
QY 112 TSEQAKTNSDKIITTSKAVNRLATGKFPVANNNTAHPKTYVD-----NTTTDTTVDRAATTERRATGARRGP 166
DB 117 TDRAATT--NDTTATNDTAVDTN-----NTTTDTTVDRAATTERRATGARRGP 166
QY 156 ----KIVPKPKIGKQLKQSSLSQDDIALGNV-----KNIRKNGKYYIYKEDGTQLQN 206
DB 167 TCGRRATPVNGNTNANNNTVTVVNDLPATNVVTDGPGSHIKTINGKQYIYVEDDGTIRKN 226
QY 207 YAL-NINGKTPFPFDETGALSN-----NTLPSKKNITNNDNTNSFAQYNQVY 252
DB 227 YVLERIGSQYFNAETGELSNOKEVRFKNGGTGSSADSTNTVTVNGDKNAF-----Y 280
QY 253 STDVANFEHVDHYLTAESWYRKYILKDGKTWTQSTEKDFRELLMTWPDQETQRYVNY 312
DB 281 GTTDKDIELVDGYFTANTWYRKEILKDGKEWTASTENDKRPILLTWWPESKAIQASLYNY 340
QY 313 MNAQ-LGIHOTNTATSPLOLNAAQTQTKIEKITAENKTNWLRQTISAFVKTSQAWN 371
DB 341 MKEOGLGNTQYTFSSSQTMQDAALEVQKRIEERAREGNTDLRTTIKNFVKTPQGMN 400
QY 372 SDEK-PFDDHLQKGLLYSNNSKLTQANSNRYLNRTPNTQTKGKOPRYTADRTIGY 430
DB 401 STSENLDDNDHLQGGALLYNNDSR-TSHANSYRLNLRTPPTSQTGKHNPKYTKDTSNGGF 459
QY 431 EFLLANDVNSPVVQAOLNWLHLMFNFGNIYANDPDANFDSIRVDVANDVNDADLLQIA 490
DB 460 EFLLANDIDNSNPAVQAOLNWLHYIMNIGTITGSEDENFDGVRVDAVDVNDVADLLQIA 519
QY 491 GYLLKAAKGIHKNKKAANDHLSILEAWSYNDDPYLHDDGDNMNMNDRRLRLSLYSLAKP 550
DB 520 SDYFRKTYGADQSQQAIAKHLISLEAWSHNDAYYNEDTKGAQLPMDPMDHMLALVYSLLRP 579
```

Search completed: February 11, 2006, 19:38:13
Job time : 191.535 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 18:56:56 ; Search time 142.338 Seconds
(without alignments)
4414.224 Million cell updates/sec

Title: US-10-797-821-36
Perfect score: 7462
Sequence: 1 METKRYXMKVKGHWTVTA.....EGKQVKGVDVAYDERLLVYR 1430

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 43937871 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*
9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than, or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7462	100.0	1430	5	Aau98029 S. mutans
2	7462	100.0	1430	7	ADD93656 Streptoco
3	7462	100.0	1430	9	ADx37279 Streptoco
4	7457	99.9	1430	5	Aau98043 S. mutans
5	7456	99.9	1430	5	Aau98042 S. mutans
6	7456	99.9	1430	5	Aau98041 S. mutans
7	7451	99.9	1430	5	Aau98044 S. mutans
8	7451	99.9	1430	5	Aau98045 S. mutans
9	3869	51.8	1476	5	Aau79284 Streptoco
10	3867	51.8	1499	7	ADC54806 Protein S
11	3845.5	51.5	1475	5	Aau98038 S. mutans
12	3835.5	51.4	1475	5	Aau98037 S. mutans
13	3829.5	51.3	1475	5	Aau98036 S. mutans
14	3828.5	51.3	1475	5	Aau98035 S. mutans
15	3828.5	51.3	1475	5	Aau98034 S. mutans
16	3823.5	51.2	1475	5	Aau98033 S. mutans
17	3823.5	51.2	1475	5	Aau98032 S. mutans
18	3822.5	51.2	1475	5	Aau98031 S. mutans
19	3821.5	51.2	1475	5	Aau98040 S. mutans
20	3818.5	51.2	1475	5	Aau98030 S. mutans
21	3817.5	51.2	1475	5	Aau98027 S. mutans
22	3817.5	51.2	1475	7	ADD93654 Streptoco
23	3817.5	51.2	1475	9	ADx37277 Streptoco
24	3790.5	50.8	1475	5	Aau98039 S. mutans

25	3788.5	50.8	1577	2	AAR91047 Alpha-D-g
26	3615	48.4	1554	7	ADD93658 Streptoco
27	3615	48.4	1554	9	ADx37281 Streptoco
28	3591.5	48.1	1375	5	Aau98028 S. mutans
29	3591.5	48.1	1375	5	Aau79288 Streptoco
30	3591.5	48.1	1375	5	ADD93655 Streptoco
31	3591.5	48.1	1375	9	ADx37278 Streptoco
32	3575	47.9	1590	9	ADD93657 Streptoco
33	3575	47.9	1590	9	ADx37280 Streptoco
34	3542	47.5	1592	2	AAR32925 Glucoseylt
35	3502.5	46.9	1518	7	ADD93660 Streptoco
36	3502.5	46.9	1518	9	ADx37283 Streptoco
37	3364.5	45.1	1497	6	ABR63234 Glucansuc
38	3334.5	44.7	2835	5	ABR98574 Amino aci
39	3334.5	44.7	2835	6	ABR55594 Amino aci
40	3227	43.2	1477	9	ADY72696 Mutant de
41	3226	43.2	1477	9	ADY72733 Mutant de
42	3224	43.2	1477	9	ADY72732 Mutant de
43	3186.5	42.7	1527	5	Aau80055 Leuconost
44	3180.5	42.6	1527	7	ADC54807 Leuconost
45	3174	42.5	1365	7	ADD93659 Streptoco

ALIGNMENTS

RESULT 1
AAU98029
ID AAU98029 standard; protein; 1430 AA.
XX AC AAU98029;
XX DT 27-AUG-2002 (first entry)
XX DE S. mutans glucosyltransferase GTFD.
XX KW Glucosyltransferase; GTFD; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
XX amyloplast; vacuole; paper manufacture.
XX OS Streptococcus mutans.
XX PN US2002031826-A1.
XX PD 14-MAR-2002.
XX PF 19-DEC-2000; 2000US-00740274.
XX PR 07-JUN-1995; 95US-00478704.
PR 07-JUN-1995; 95US-00482711.
PR 07-JUN-1995; 95US-00485243.
PR 16-JAN-1998; 98US-00007999.
PR 16-JAN-1998; 98US-00008172.
PR 20-JAN-1998; 98US-00009620.
PR 11-DEC-1998; 98US-00210361.
(NICH//) NICHOLS S E.
XX Nichols SE;
XX WPI; 2002-414332/44.
XX N-PSDB; ABK52940.
XX Glucosyltransferase B or D protein useful for producing a glucan useful
as substitutes for and additions to modified starch and latexes in paper
manufacture, comprises mutations in specific positions.
XX Disclosure; Page 38-42; 44pp; English.
XX The invention an isolated protein comprising a glucosyltransferase (GTF)
B polypeptide having changes at position from 1448V, D457N, D567T,
CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T,
CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a

CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
CC complementary polynucleotide, a ribonucleic acid sequence encoding the
CC GTF mutant, an expression cassette comprising the polynucleotide operably
CC linked to a promoter, a vector comprising the expression cassette, host
CC cell introduced with the vector, a transgenic plant comprising the
CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
CC coating composition comprising a glucan produced in a plant transformed
CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
CC thermoplastic molecule or their combinations or glucan and starch where
CC the glucan is produced in the amyloplast and/or vacuole or a maize line
CC deficient in starch biosynthesis, transformed with a gene encoding a
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
CC comprising the glucan (paper sizing/coating agent). The vector is useful
CC for producing a glucan in a plant. The method comprises transforming a
CC plant cell with the vector, growing the plant cell under plant growing
CC conditions to produce a regenerated plant and inducing expression of the
CC polynucleotide for a time sufficient to produce the glucan in the
CC regenerated plant, where the vector contains a transit sequence from
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
CC is produced in the amyloplast of potato or the vacuole of sugar beet.
CC Glucans are useful as substitutes for and additions to modified starch
CC and latexes in paper manufacture. Unlike prior art techniques, which
CC require input materials that produce chemical effluents, paper
CC manufacture utilises the glucan produced by GTF, which utilises
CC biologically produced input materials, is more cost-effective and
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
CC properties and impart gloss to the paper during coating step. The present
CC sequence represents GTFD
XX
SQ Sequence 1430 AA;

Query Match 100.0%; Score 7462; DB 5; Length 1430;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 METKRYKMHKVKHVVAVASGLITLTTTGLSSVSAETEQTSQKVVTKSDDKAA 60
DB |||||
QY 1 METKRYKMHKVKHVVAVASGLITLTTTGLSSVSAETEQTSQKVVTKSDDKAA 60
DB |||||
QY 61 SSSQTDAPKTKQAOTEQTAQOANVADTSTSTKETPSQNIITTAQNSDDKVTNTKSE 120
DB |||||
QY 61 SSSQTDAPKTKQAOTEQTAQOANVADTSTSTKETPSQNIITTAQNSDDKVTNTKSE 120
DB |||||
QY 121 EAQTSSEERTKQSEEAQTASSQALTOAKAELTKQRTAAQENKPNVDAIIPNVKQIDGK 180
DB |||||
QY 121 EAQTSSEERTKQSEEAQTASSQALTOAKAELTKQRTAAQENKPNVDAIIPNVKQIDGK 180
DB |||||
QY 181 YYYIGSDGQPKNFALTVNNKVLFPDNTGALTDTSTQYQFQKGLTKLNNDYTPHNQIVNF 240
DB |||||
QY 181 YYYIGSDGQPKNFALTVNNKVLFPDNTGALTDTSTQYQFQKGLTKLNNDYTPHNQIVNF 240
DB |||||
QY 241 ENTSLETIDNVYTDSTWYRPDKILKNGKTWTASSSDRLPLMSWPKQTOIAIYLNMYN 300
DB |||||
QY 241 ENTSLETIDNVYTDSTWYRPDKILKNGKTWTASSSDRLPLMSWPKQTOIAIYLNMYN 300
DB |||||
QY 301 QOGLGTGENYTDSSQESLNLAQTVQVKIETKISQTOQWLRDIINSFVKTPQNWNSQ 360
DB |||||
QY 301 QOGLGTGENYTDSSQESLNLAQTVQVKIETKISQTOQWLRDIINSFVKTPQNWNSQ 360
DB |||||
QY 361 TESDTSAGEKHQLOGGALLYSNDKTVANSDYRLNRTPTSTQTKPKYFEDNSSGGYDF 420
DB |||||
QY 361 TESDTSAGEKHQLOGGALLYSNDKTVANSDYRLNRTPTSTQTKPKYFEDNSSGGYDF 420
DB |||||
QY 421 LLANDIDNSNPVQAEQLNWLHYLMNYGSIIVANDPEANPDGVRVDVANNVADLLQIASD 480
DB |||||
QY 421 LLANDIDNSNPVQAEQLNWLHYLMNYGSIIVANDPEANPDGVRVDVANNVADLLQIASD 480
DB |||||
QY 481 YLKAHYGVDKSEKNAINHLSILEAWSNDNDPQNKDTKGAQLPIDNKLRLSLLYALTRPLE 540
DB |||||
QY 481 YLKAHYGVDKSEKNAINHLSILEAWSNDNDPQNKDTKGAQLPIDNKLRLSLLYALTRPLE 540
DB |||||

RESULT 2
ADD93656
ID ADD93656 standard; protein; 1430 AA.
XX
AC ADD93656;
XX
DT 29-JAN-2004 (first entry)
XX
DE Streptococcus mutans glucosyltransferase-D.
XX
KW Glucosyltransferase; enzyme; vaccine; antiscar; immunogen.

QY 541 KDAKNKNEIRSGLEBPVITNSLNRSAGKNSERMANYIFIRAHDSVQTVIAKIIKAQIN 600
DB |||||
QY 541 KDAKNKNEIRSGLEBPVITNSLNRSAGKNSERMANYIFIRAHDSVQTVIAKIIKAQIN 600
DB |||||
QY 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKKTQSNIPATAYALMLSNKDSITRLYYGDMYS 660
DB |||||
QY 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKKTQSNIPATAYALMLSNKDSITRLYYGDMYS 660
DB |||||
QY 661 DDGQYMATKSPYDAIDTLLKARIKAYAGGQDMKITYYVEGDKSHMDWDYTGVLTSVRYGT 720
DB |||||
QY 661 DDGQYMATKSPYDAIDTLLKARIKAYAGGQDMKITYYVEGDKSHMDWDYTGVLTSVRYGT 720
DB |||||
QY 721 GANEATDQGSSEATKTQGHAVITSSNPSSKLQNQDKVIYVMGAHKNQYRPLLLTTKQGL 780
DB |||||
QY 721 GANEATDQGSSEATKTQGHAVITSSNPSSKLQNQDKVIYVMGAHKNQYRPLLLTTKQGL 780
DB |||||
QY 781 TSYTSDAAAKSLYKRTNDKGBELVFPDASDIQGYLNPQVSGYLAVWVPVGSADNDQVRVAAS 840
DB |||||
QY 781 TSYTSDAAAKSLYKRTNDKGBELVFPDASDIQGYLNPQVSGYLAVWVPVGSADNDQVRVAAS 840
DB |||||
QY 841 NKANATQGVYESSGALDSQLIYEGFSNFQDFVTKDSYTNKKIAQNVOLFSGWGTSPFM 900
DB |||||
QY 841 NKANATQGVYESSGALDSQLIYEGFSNFQDFVTKDSYTNKKIAQNVOLFSGWGTSPFM 900
DB |||||
QY 901 APQVYSSDGSFLDSIIQNGYAFEDRYDLAKSKNNKYGSGQODMINAVKALHKSIGQVAD 960
DB |||||
QY 901 APQVYSSDGSFLDSIIQNGYAFEDRYDLAKSKNNKYGSGQODMINAVKALHKSIGQVAD 960
DB |||||
QY 961 WVPDQIYNLPKGVVTVATRVNDYGEYKDSSEIKNTLYAANTKSGNKDQYQAKYGGAFLSEL 1020
DB |||||
QY 961 WVPDQIYNLPKGVVTVATRVNDYGEYKDSSEIKNTLYAANTKSGNKDQYQAKYGGAFLSEL 1020
DB |||||
QY 1021 AAKYPSIFENRTQISNGKKIDPSEKITAMKAKYFNGTNIILGRGVGYVLKDNASDKYFELKG 1080
DB |||||
QY 1021 AAKYPSIFENRTQISNGKKIDPSEKITAMKAKYFNGTNIILGRGVGYVLKDNASDKYFELKG 1080
DB |||||
QY 1081 NQTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYFDDNHHMYVGL 1140
DB |||||
QY 1081 NQTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYFDDNHHMYVGL 1140
DB |||||
QY 1141 QQLNGEVQYFLSNGVQLRESFLENADGSKNYPHGLGNYSNGYYSFDDNDSKWRYFDASGV 1200
DB |||||
QY 1141 QQLNGEVQYFLSNGVQLRESFLENADGSKNYPHGLGNYSNGYYSFDDNDSKWRYFDASGV 1200
DB |||||
QY 1201 MAVGLKTINGNTQYFDQDGYQVKGAWITGSDGKKRYFDDGSGNMAVNRFANDKNGDWWYL 1260
DB |||||
QY 1201 MAVGLKTINGNTQYFDQDGYQVKGAWITGSDGKKRYFDDGSGNMAVNRFANDKNGDWWYL 1260
DB |||||
QY 1261 NSDGIALVGVOTINGKTYTFQDQKQIKGKIITDNGKLYFLANSSELARNIFATDSQNN 1320
DB |||||
QY 1261 NSDGIALVGVOTINGKTYTFQDQKQIKGKIITDNGKLYFLANSSELARNIFATDSQNN 1320
DB |||||
QY 1321 WYFSGDGVAVTGSQTIAGKKLYPASDCKQVKGSPVYNGKHYHYHADSSELQVNRFEAD 1380
DB |||||
QY 1321 WYFSGDGVAVTGSQTIAGKKLYPASDCKQVKGSPVYNGKHYHYHADSSELQVNRFEAD 1380
DB |||||
QY 1381 KDGNNYLLDSNGEALTGSGRINDQVRVFTREGKQVKGVDVADERLLLYR 1430
DB |||||
QY 1381 KDGNNYLLDSNGEALTGSGRINDQVRVFTREGKQVKGVDVADERLLLYR 1430
DB |||||

XX OS Streptococcus mutans.
XX PN WO2003075845-A2.
XX PD 18-SEP-2003.
XX PF 07-MAR-2003; 2003WO-US006962.
XX PR 07-MAR-2002; 2002US-0363209P.
XX PR 08-AUG-2002; 2002US-0402483P.
XX PA (FORS-) FORSYTH INST.
XX PI Smith DJ, Taubman MA;
XX DR WPI; 2003-845091/78.
XX PT Composition useful as vaccines for dental caries comprises a fragment of
XX PT a glucan binding protein-B binding to a major histocompatibility complex
XX PT class II protein.
XX PS Claim 16; Page 13-14; 49pp; English.
XX CC The present sequence is the protein sequence of Streptococcus mutans
XX CC glucosyltransferase-D (GTF-D). Peptide fragments of GTF-D, especially
XX CC from the catalytic domain of the polypeptide, can be used in immunogenic
XX CC compositions and subunit vaccines for dental caries. These compositions
XX CC comprise a major histocompatibility complex (MHC) class II protein-
XX CC binding peptide from S. mutans glucan binding protein-B (GbpB)
XX CC covalently linked with a peptide fragment of a streptococcal
XX CC glucosyltransferase. The compositions are used in a claimed method of
XX CC eliciting production of an antibody in a mammal. Diepitopic or
XX CC multipitopic polypeptides can be prepared synthetically or by
XX CC recombinant DNA technology. Antibodies raised against MHC class II
XX CC binding fragments of GbpB can be used in passive immunisation.
XX SQ Sequence 1430 AA;
Query Match 100.0%; Score 7462; DB 7; Length 1430;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 METKRYKMHKVKHWHVAVASGLITLTTTIGSSVSAETEQQTSDKVVTKQSEDDKAA 60
DB 1 METKRYKMHKVKHWHVAVASGLITLTTTIGSSVSAETEQQTSDKVVTKQSEDDKAA 60
QY 61 SESSQTDAPKTKQAQTEQTAQSQANVADTSTSIITKETPSQNTTQANSDDKVTNTKSE 120
DB 61 SESSQTDAPKTKQAQTEQTAQSQANVADTSTSIITKETPSQNTTQANSDDKVTNTKSE 120
QY 121 EAQTSSEERTKQSEEAQTASSQALTOAKELTKQRTAAQENKPNVDLAAIPNVKQIDGK 180
DB 121 EAQTSSEERTKQSEEAQTASSQALTOAKELTKQRTAAQENKPNVDLAAIPNVKQIDGK 180
QY 181 YYYIGSDQPKGNFALTVNNKVLVFDKNTGALTDSOYQFKGLTKLNDYTPHNOIVNF 240
DB 181 YYYIGSDQPKGNFALTVNNKVLVFDKNTGALTDSOYQFKGLTKLNDYTPHNOIVNF 240
QY 241 ENTSLTIDNVYVADSWYRPRKIDLNGKKTWTASSSEDLRLPLMSWPKQTOIAYLNTMN 300
DB 241 ENTSLTIDNVYVADSWYRPRKIDLNGKKTWTASSSEDLRLPLMSWPKQTOIAYLNTMN 300
QY 301 QOGLGTGENYADSSQESLNLAQTVQVKIETKISQTOQTQWLRIINSFVKTPQNNWSQ 360
DB 301 QOGLGTGENYADSSQESLNLAQTVQVKIETKISQTOQTQWLRIINSFVKTPQNNWSQ 360
QY 361 TESDTSAGEKDHQLOGGALLYNSDKTAVANDSYRLNRTPTSQTKPKYFEDNSSGGYDF 420
DB 361 TESDTSAGEKDHQLOGGALLYNSDKTAVANDSYRLNRTPTSQTKPKYFEDNSSGGYDF 420
QY 421 LLANDIDNSNPVVOAEQLNLHLYLMNYGSIVANDPEANFDGVRVDVNDVNADLLQIASD 480

DB 421 LLANDIDNSNPVVOAEQLNLHLYLMNYGSIVANDPEANFDGVRVDVNDVNADLLQIASD 480
QY 481 YLKAHYGVDDKSEKNAIHLISILEANDNDPOYNKDTKGALPLDNKLRLLSLVALTRPLE 540
DB 481 YLKAHYGVDDKSEKNAIHLISILEANDNDPOYNKDTKGALPLDNKLRLLSLVALTRPLE 540
QY 541 KDAENKNEIRSGLEPVITNSLNNRSGKNSERMANYIFIRAHDSVQTVIAKIIKAQIN 600
DB 541 KDAENKNEIRSGLEPVITNSLNNRSGKNSERMANYIFIRAHDSVQTVIAKIIKAQIN 600
QY 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYYTQSNIPATAYALMLSNKDSITRLYYGDMYS 660
DB 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYYTQSNIPATAYALMLSNKDSITRLYYGDMYS 660
QY 661 DDQOYMATKSPYYDAIDTLKARIKYAAGSQDMKITYVEGDKSHMDWDVTGVLTSVRYGT 720
DB 661 DDQOYMATKSPYYDAIDTLKARIKYAAGSQDMKITYVEGDKSHMDWDVTGVLTSVRYGT 720
QY 721 GANEATDQSEATKTQGMVITSNPNPSLKLNDKQVIVNMGAHKNQYRPLLLTTKDG 780
DB 721 GANEATDQSEATKTQGMVITSNPNPSLKLNDKQVIVNMGAHKNQYRPLLLTTKDG 780
QY 781 TSYTSDAAKSLYRKTNDKGLVFDASDIQGYLNPQVSGYLAWVPVGASDNDQVRVAAS 840
DB 781 TSYTSDAAKSLYRKTNDKGLVFDASDIQGYLNPQVSGYLAWVPVGASDNDQVRVAAS 840
QY 841 NKANATGOVYESSALDSQLIYEGFSNFODFTVKDSYTNKKIAQNVQLFKSHGVTSEF 900
DB 841 NKANATGOVYESSALDSQLIYEGFSNFODFTVKDSYTNKKIAQNVQLFKSHGVTSEF 900
QY 901 APQYVSSSEDSGLDSIIQNGYAFEDRYDLAMSKNNKYGSSQDMINAVKALHSGIOVIAD 960
DB 901 APQYVSSSEDSGLDSIIQNGYAFEDRYDLAMSKNNKYGSSQDMINAVKALHSGIOVIAD 960
QY 961 WYVDQIYNLPGKEVVTATRVNDYGYRKDSEIKNTLYAANTKSGKDYQAKYGGAFSEL 1020
DB 961 WYVDQIYNLPGKEVVTATRVNDYGYRKDSEIKNTLYAANTKSGKDYQAKYGGAFSEL 1020
QY 1021 AAKYPSIFNRTQISNGKKIDPSEKITAMKAKYFNGTNIILGRGVYVLDKNSDKYFELKG 1080
DB 1021 AAKYPSIFNRTQISNGKKIDPSEKITAMKAKYFNGTNIILGRGVYVLDKNSDKYFELKG 1080
QY 1081 NQYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYFDNNGHMYGL 1140
DB 1081 NQYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYFDNNGHMYGL 1140
QY 1141 QQLNGEVQYFLSNGVOLRESFLENADGSKNYFGLGNRYSGYYSFDNDSKWRYFDASGV 1200
DB 1141 QQLNGEVQYFLSNGVOLRESFLENADGSKNYFGLGNRYSGYYSFDNDSKWRYFDASGV 1200
QY 1201 MAVGLKTTINGTOYFDQDGYQVKGAWITGSDGKKYFDDGSGNMAVNRFANDKNGDWWYL 1260
DB 1201 MAVGLKTTINGTOYFDQDGYQVKGAWITGSDGKKYFDDGSGNMAVNRFANDKNGDWWYL 1260
QY 1261 NSDGLALGVQVOTINGKTYFQDQYQVKGAWITGSDGKKYFDDGSGNMAVNRFANDKNGDWWYL 1320
DB 1261 NSDGLALGVQVOTINGKTYFQDQYQVKGAWITGSDGKKYFDDGSGNMAVNRFANDKNGDWWYL 1320
QY 1321 WYFSGDGVAVTGSQTIAGKKLYFASDGKQVKGSPVYNGKVHYHADSGELQVNRFEAD 1380
DB 1321 WYFSGDGVAVTGSQTIAGKKLYFASDGKQVKGSPVYNGKVHYHADSGELQVNRFEAD 1380
QY 1381 KQGNWYVLDNSGEALTGSRINDQVFFTRREGQVKGVDVDERLLVYR 1430
DB 1381 KQGNWYVLDNSGEALTGSRINDQVFFTRREGQVKGVDVDERLLVYR 1430
RESULT 3
ADX37279
ID ADX37279 standard; protein; 1430 AA.
XX
AC ADX37279;
XX

DT 21-APR-2005 (first entry)
 XX Streptococcus mutant glucan binding protein B variant #8.
 DE immunogenicity; immune stimulation; glucan binding protein-B;
 KW microparticle; major histocompatibility complex; tooth disease.
 XX Streptococcus mutans.
 OS US2005031633-A1.
 FN 10-FEB-2005.
 PD 09-MAR-2004; 2004US-00797821.
 PF 13-APR-1998; 98US-0081550P.
 XX 08-JAN-1999; 99US-0115142P.
 PR 12-APR-1999; 99US-00290049.
 PR 07-MAR-2002; 2002US-0363209P.
 PR 08-AUG-2002; 2002US-0402483P.
 PR 07-MAR-2003; 2003US-00383930.
 XX (SMIT/) SMITH J J.
 PA (TAUB/) TAUBMAN M A.
 PA Smith DJ, Taubman MA;
 PI WPI; 2005-151644/16.
 DR
 XX New composition comprising a fragment of a glucan binding protein-B (GbpB) that binds to MHC class II protein, and a biocompatible microparticle, useful for producing an antibody (claimed) for immunizing mammals against dental caries.
 PT
 PT Claim 7; SEQ ID NO 36; 73pp; English.
 XX
 PS The invention relates to a composition comprising a fragment of a glucan binding protein-B (GbpB) and a biocompatible microparticle, where the fragment binds to a major histocompatibility complex (MHC) class II protein. The composition is useful for producing an antibody for immunizing mammals against dental caries. This sequence corresponds to a Streptococcus mutans GbpB protein of the invention.
 CC
 CC Sequence 1430 AA;
 SQ
 Query Match 100.0%; Score 7462; DB 9; Length 1430;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 METKRYKMKYKHKHVVAVASGLITLTGTLGSSVSAETEQOTSDDKVVTQKSEDDKAA 60
 DB 1 METKRYKMKYKHKHVVAVASGLITLTGTLGSSVSAETEQOTSDDKVVTQKSEDDKAA 60
 QY 61 SESSOTDAPKTKQAOTQEQTAQSQANVADTSTSTKETPSQNIITQANSDDKTIVNTKSE 120
 DB 61 SESSOTDAPKTKQAOTQEQTAQSQANVADTSTSTKETPSQNIITQANSDDKTIVNTKSE 120
 QY 121 EAQTSERTKQSEEAQTASSQALTOAKAELTKQRTAAQENKPNVDLAAIPNVKQIDGK 180
 DB 121 EAQTSERTKQSEEAQTASSQALTOAKAELTKQRTAAQENKPNVDLAAIPNVKQIDGK 180
 QY 181 YYYIGSDGQPKNFALTNNKVLVYFDKNTGALTDTSTQYQKGLTKLNDYTPHNIQVNF 240
 DB 181 YYYIGSDGQPKNFALTNNKVLVYFDKNTGALTDTSTQYQKGLTKLNDYTPHNIQVNF 240
 QY 241 ENTSLETIDNYTADSWYRPKDIKNGKKTWTASSSDLRPLLMSWMPDKQTOIAYLNYN 300
 DB 241 ENTSLETIDNYTADSWYRPKDIKNGKKTWTASSSDLRPLLMSWMPDKQTOIAYLNYN 300
 QY 301 QOGLGTGENYTADSSQESLNLAQTVQVKIETKISQTOQTOWLRDIINSFVKTPQNNWSQ 360
 DB 301 QOGLGTGENYTADSSQESLNLAQTVQVKIETKISQTOQTOWLRDIINSFVKTPQNNWSQ 360

QY 361 TESDTSAAGEKDLQGGALLYNSDKTAYANSYRLNRTPTSTQTKPKYFEDNSSGGYDF 420
 DB 361 TESDTSAAGEKDLQGGALLYNSDKTAYANSYRLNRTPTSTQTKPKYFEDNSSGGYDF 420
 QY 421 LLANDIDNSNPVQAEQLNWLHLNYGSIIVANDPEANFPGVVRVDAVNVNADLLQIASD 480
 DB 421 LLANDIDNSNPVQAEQLNWLHLNYGSIIVANDPEANFPGVVRVDAVNVNADLLQIASD 480
 QY 481 YLKAHYGVDSKSEKNAINHLSILEAWSNDNDPOYKDTKGAQLPIDNKLRLSLLYALTRPLE 540
 DB 481 YLKAHYGVDSKSEKNAINHLSILEAWSNDNDPOYKDTKGAQLPIDNKLRLSLLYALTRPLE 540
 QY 541 KDAENKNEIRSGLEPVIITNSLNNRSEAGKNSERMANYIFIRAHDEVOVIAKIIKAQIN 600
 DB 541 KDAENKNEIRSGLEPVIITNSLNNRSEAGKNSERMANYIFIRAHDEVOVIAKIIKAQIN 600
 QY 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKKTQSNIPTAYALMLSNKDSITRLYIGDMYS 660
 DB 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKKTQSNIPTAYALMLSNKDSITRLYIGDMYS 660
 QY 661 DDGQYMATKSPYYDAIDTLKARIKYAAAGQDMKITYVEGDKSHMDWDYTGVLTSVRYGT 720
 DB 661 DDGQYMATKSPYYDAIDTLKARIKYAAAGQDMKITYVEGDKSHMDWDYTGVLTSVRYGT 720
 QY 721 GANEATDQGSBATKTQGMAVITSSNPSLKLQNDKVIIVNMGAHKNQOEYRPLLLTTKQGL 780
 DB 721 GANEATDQGSBATKTQGMAVITSSNPSLKLQNDKVIIVNMGAHKNQOEYRPLLLTTKQGL 780
 QY 781 TSYTSDAAAKSLYRKNTDKGELVFPDASDIQGYLNPQVSGYLAVWPVPGASDNQDVRVAAS 840
 DB 781 TSYTSDAAAKSLYRKNTDKGELVFPDASDIQGYLNPQVSGYLAVWPVPGASDNQDVRVAAS 840
 QY 841 NKANATQGVYESSSALDSQLIYEGFSNFDQFVTKDSYTNKKIAQNVQLFKSWGVTSPFM 900
 DB 841 NKANATQGVYESSSALDSQLIYEGFSNFDQFVTKDSYTNKKIAQNVQLFKSWGVTSPFM 900
 QY 901 APOYVSSSDGSLDSIIQNGVAFEDRYDLAMSNNKYGSGQDDMINAVKALHKSGLQVAD 960
 DB 901 APOYVSSSDGSLDSIIQNGVAFEDRYDLAMSNNKYGSGQDDMINAVKALHKSGLQVAD 960
 QY 961 WVPDQIYNLPGKEVVVTAIRVNDYGEYRKDSEIKNTLYAANTKSNKDKYQAKYGGAFLSEL 1020
 DB 961 WVPDQIYNLPGKEVVVTAIRVNDYGEYRKDSEIKNTLYAANTKSNKDKYQAKYGGAFLSEL 1020
 QY 1021 AAKYPSIIFNRTOISNGKKIDPSEKITAWKAKYFNGTNIILGRGVGVYLDKNASDKYFELKG 1080
 DB 1021 AAKYPSIIFNRTOISNGKKIDPSEKITAWKAKYFNGTNIILGRGVGVYLDKNASDKYFELKG 1080
 QY 1081 NOTYLPKQMTNKEASTGFVNDGNGMTFYTSQYQAKNSFVODAKGNWYTFDNNGHMVYGL 1140
 DB 1081 NOTYLPKQMTNKEASTGFVNDGNGMTFYTSQYQAKNSFVODAKGNWYTFDNNGHMVYGL 1140
 QY 1141 QQLANGEVOYFLSNGVQLRESFLENADGSKNYPFGLHNGRYSNGYYSFNDNSKWRYFDASGV 1200
 DB 1141 QQLANGEVOYFLSNGVQLRESFLENADGSKNYPFGLHNGRYSNGYYSFNDNSKWRYFDASGV 1200
 QY 1201 MAVGLKTINGNTQYPDQGYQVKGAWITGSDCKKRYFDDGSGMAVNRFPANDKNGDWYVL 1260
 DB 1201 MAVGLKTINGNTQYPDQGYQVKGAWITGSDCKKRYFDDGSGMAVNRFPANDKNGDWYVL 1260
 QY 1261 NSDGTALVGVTINGKTYTFGDDGKQIKGKIITDNGKLKYFLANSSELARNIPATDSQNN 1320
 DB 1261 NSDGTALVGVTINGKTYTFGDDGKQIKGKIITDNGKLKYFLANSSELARNIPATDSQNN 1320
 QY 1321 WYFSGDGVAVTGSQTIAGKKLYFASDGKQVKGSEFVYNGKHYHYHADSSELQVNRFEAD 1380
 DB 1321 WYFSGDGVAVTGSQTIAGKKLYFASDGKQVKGSEFVYNGKHYHYHADSSELQVNRFEAD 1380
 QY 1381 KDGNNYILDSNGEALTGSQRINDQVFFTRREGQVKGDAVADERLLVYR 1430
 DB 1381 KDGNNYILDSNGEALTGSQRINDQVFFTRREGQVKGDAVADERLLVYR 1430

RESULT 4
AAU98043
ID AAU98043 standard; protein; 1430 AA.
XX
AC AAU98043;
XX
DT 27-AUG-2002 (first entry)
XX
DE S. mutans glucosyltransferase GTFD mutant N471D.
XX
KW Glucosyltransferase; GTFD; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture; mutant; mutein.
XX
OS Streptococcus mutans.
OS Synthetic.
XX
XX
XX Key Location/Qualifiers
FH Misc-difference 471 /note= "Wild-type Asn substituted by Asp"
FT
XX US2002031826-A1.
XX
XX 14-MAR-2002.
XX
XX 19-DEC-2000; 2000US-00740274.
XX
XX 07-JUN-1995; 95US-00478704.
XX 07-JUN-1995; 95US-00482711.
XX 07-JUN-1995; 95US-00485243.
XX 16-JAN-1998; 98US-00007999.
XX 16-JAN-1998; 98US-00008172.
XX 20-JAN-1998; 98US-00009620.
XX 11-DEC-1998; 98US-00210361.
XX
XX (NICH/) NICHOLS S E.
XX
XX Nichols SE; |
XX
XX WPI; 2002-414332/44.
XX
XX Glucosyltransferase B or D protein useful for producing a glucan useful
XX as substitutes for and additions to modified starch and latexes in paper
XX manufacture, comprises mutations in specific positions.
XX
XX Claim 36; Page; 44pp; English.
XX
XX The invention an isolated protein comprising a glucosyltransferase (GTF)
XX B polypeptide having changes at position from 1448V, D457N, D567T,
XX K1014T, D457N/D567T, D457N/D571K, D567T/D571K/D1014T,
XX I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
XX GTF D polypeptide having changes at positions from T589D, T589E, N471D,
XX N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
XX GTF mutant, an isolated polynucleotide which encodes F1 or P2, or its
XX complementary polynucleotide, a ribonucleic acid sequence encoding the
XX GTF mutant, an expression cassette comprising the polynucleotide operably
XX linked to a promoter, a vector comprising the expression cassette, host
XX cell introduced with the vector, a transgenic plant comprising the
XX vector, a seed or tuber from the transgenic plant, a paper sizing and/or
XX coating composition comprising a glucan produced in a plant transformed
XX with a gene encoding the mutant GTF, wild type or, starch, a latex,
XX thermoplastic molecule or their combinations or glucan and starch where
XX the glucan is produced in the amyloplast and/or vacuole or a maize line
XX deficient in starch biosynthesis, transformed with a gene encoding a
XX glucosyltransferase B or D enzyme, wild-type or mutant and a paper
XX comprising the glucan (paper sizing/coating agent). The vector is useful
XX for producing a glucan in a plant. The method comprises transforming a
XX plant cell with the vector, growing the plant cell under plant growing
XX conditions to produce a regenerated plant and inducing expression of the
XX polynucleotide for a time sufficient to produce the glucan in the
XX regenerated plant, where the vector contains a transit sequence from
XX ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
XX chlorophyll AB binding protein to produce a transgenic plant, and glucan

CC is produced in the amyloplast of potato or the vacuole of sugar beet.
CC Glucans are useful as substitutes for and additions to modified starch
CC and latexes in paper manufacture. Unlike prior art techniques, which
CC require input materials that produce chemical effluents, paper
CC manufacture utilising the glucan produced by GTF, which utilises
CC biologically produced input materials, is more cost-effective and
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
CC properties and impart gloss to the paper during coating step. The present
CC sequence represents a GTFD mutant of the invention. Note: The present
CC sequence is not shown in the specification but was created by the indexer
CC using the GTFD sequence appearing as AAU98029 and the information in
CC claim 36
XX
SQ Sequence 1430 AA;
Query Match 99.9%; Score 7457; DB 5; Length 1430;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1429; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 METKRRYKMKHKVKGHWVTVAVASGLITLTGTTTILGSSVSAETEQTSDKVVTKSEDDKAA 60
DB 1 METKRRYKMKHKVKGHWVTVAVASGLITLTGTTTILGSSVSAETEQTSDKVVTKSEDDKAA 60
QY 61 SSSQTDAPKTKQAQTEQTQAQSQANVADTSTSTITKETPSQNTTQANSDDKTVTNTKSE 120
DB 61 SSSQTDAPKTKQAQTEQTQAQSQANVADTSTSTITKETPSQNTTQANSDDKTVTNTKSE 120
QY 121 EAQTSSEERTKQSEEAQTASSOALTOAKAELTKOROTAAQENKNPVDLAAIPNVKQIDGK 180
DB 121 EAQTSSEERTKQSEEAQTASSOALTOAKAELTKOROTAAQENKNPVDLAAIPNVKQIDGK 180
QY 181 YYYIGSDGQPKKFALTNNKVLVFDKNTGALTDTTSQYQFKQGLTKLNNDYTPHNQIVNF 240
DB 181 YYYIGSDGQPKKFALTNNKVLVFDKNTGALTDTTSQYQFKQGLTKLNNDYTPHNQIVNF 240
QY 241 ENTSLETIDNYVTADSWTRPKDILKNGKTTWTASSEDRLPPLMSWMPDKQTQIAYLNYN 300
DB 241 ENTSLETIDNYVTADSWTRPKDILKNGKTTWTASSEDRLPPLMSWMPDKQTQIAYLNYN 300
QY 301 QQGLGTGENYTDSSQESLNLAQTVOVKIETKISQTOQTOWLRDILINSFVKTOPNNSQ 360
DB 301 QQGLGTGENYTDSSQESLNLAQTVOVKIETKISQTOQTOWLRDILINSFVKTOPNNSQ 360
QY 361 TESDTSAGEKHLOGGALLYNSDKTAVANSYRLLNRPPTSQTGPKPKYFEDNSSGGYDF 420
DB 361 TESDTSAGEKHLOGGALLYNSDKTAVANSYRLLNRPPTSQTGPKPKYFEDNSSGGYDF 420
QY 421 LLANDIDNSNPVQAEQLNHLHLYLNYGSIIVANDPEANPDGVRVDAVDNVDADLLQIASD 480
DB 421 LLANDIDNSNPVQAEQLNHLHLYLNYGSIIVANDPEANPDGVRVDAVDNVDADLLQIASD 480
QY 481 YLKAHYGVDKSEKNAIHLSTILEAWSNDNDPQYNKDTKGALPIDNKLRLSLLYALTRPLE 540
DB 481 YLKAHYGVDKSEKNAIHLSTILEAWSNDNDPQYNKDTKGALPIDNKLRLSLLYALTRPLE 540
QY 541 KDASNKNIRSGLEPVIITNSLNRSAGKNSRMANYIFIRAHDSVQTVIAKIIKAQIN 600
DB 541 KDASNKNIRSGLEPVIITNSLNRSAGKNSRMANYIFIRAHDSVQTVIAKIIKAQIN 600
QY 601 PKTDGLTFTLDELKQAFKIYNEQDMROAKKCYTQSNIPATAYALMLSNKOSITRLYYGDMYS 660
DB 601 PKTDGLTFTLDELKQAFKIYNEQDMROAKKCYTQSNIPATAYALMLSNKOSITRLYYGDMYS 660
QY 661 DDGQYMATKSPYDAIDTLLKARIKYAAGQDMKITTYEGDKSHMDWDYTGVLTSVRYGT 720
DB 661 DDGQYMATKSPYDAIDTLLKARIKYAAGQDMKITTYEGDKSHMDWDYTGVLTSVRYGT 720
QY 721 GANEATDQGSSEATKTQGMVITSSNPNKLQNDKVIIVNMGAHKNQEVRLPILLTTKQGL 780
DB 721 GANEATDQGSSEATKTQGMVITSSNPNKLQNDKVIIVNMGAHKNQEVRLPILLTTKQGL 780
QY 781 TSYTSDAAAKSLYRKNTDKGELVDFDSDIQGLNPQVSGYLAVWVPVPGASDQDVRVAAS 840
DB 781 TSYTSDAAAKSLYRKNTDKGELVDFDSDIQGLNPQVSGYLAVWVPVPGASDQDVRVAAS 840

Db 781 TSYTSDAAKSLYRKTNKGELVFDASDIQGYLNPQVSGYLAVVPVPGASDNQDVRVAAS 840
QY 841 NKANATGQVYESSSALDSQLIYEGFSNFQDFTVKDSYTNKKIAQNVQLFKSGWVTSPEM 900
Db 841 NKANATGQVYESSSALDSQLIYEGFSNFQDFTVKDSYTNKKIAQNVQLFKSGWVTSPEM 900
QY 901 APQYVSSDGSFLDSIQNGYAFEDRYDLAMSKNNKYGSOQDMINAVKALHKSGLQVAD 960
Db 901 APQYVSSDGSFLDSIQNGYAFEDRYDLAMSKNNKYGSOQDMINAVKALHKSGLQVAD 960
QY 961 WYPDOIYNLPKGVEVTVATRVNDYGYRKDSEIKNTLYAANTKSNKGYQAKYGGAFLSL 1020
Db 961 WYPDOIYNLPKGVEVTVATRVNDYGYRKDSEIKNTLYAANTKSNKGYQAKYGGAFLSL 1020
QY 1021 AAKYPSIFNRTQISNGKKIDPSEKITAWEKAKYFNGTILGRGVGYVLKDNASDKYFELKG 1080
Db 1021 AAKYPSIFNRTQISNGKKIDPSEKITAWEKAKYFNGTILGRGVGYVLKDNASDKYFELKG 1080
QY 1081 NOTYLPKQNTKEASTGFVNDGNGMTFYSTGYQAKNSFVQDAKGNWYFEDNNGHMYGL 1140
Db 1081 NOTYLPKQNTKEASTGFVNDGNGMTFYSTGYQAKNSFVQDAKGNWYFEDNNGHMYGL 1140
QY 1141 QQLNGEVQYFLSNGVQLRESFLENADGSKNYFGLNRYSGYYSFNDNSKWRYPDASGV 1200
Db 1141 QQLNGEVQYFLSNGVQLRESFLENADGSKNYFGLNRYSGYYSFNDNSKWRYPDASGV 1200
QY 1201 MAVGLKTINGNTQYFDQGYQVKGAWITGSDGKKRYFDDGSGNMAVNFANDKNGDWYVL 1260
Db 1201 MAVGLKTINGNTQYFDQGYQVKGAWITGSDGKKRYFDDGSGNMAVNFANDKNGDWYVL 1260
QY 1261 NSDGTALVCVORTNGKTYFVGQDGQIKGKIITDNGKLYFLANSSELARNIFATDSQNN 1320
Db 1261 NSDGTALVCVORTNGKTYFVGQDGQIKGKIITDNGKLYFLANSSELARNIFATDSQNN 1320
QY 1321 WYFSGDGVAVTGSOTIAGKLYFASDGKQVKGSVFYNGKVHYHADSGELQVNRFEAD 1380
Db 1321 WYFSGDGVAVTGSOTIAGKLYFASDGKQVKGSVFYNGKVHYHADSGELQVNRFEAD 1380
QY 1381 KDGNNYLDNSGEALTGSGRINDQVFFTRGKQVKGVDVYDERLLVYR 1430
Db 1381 KDGNNYLDNSGEALTGSGRINDQVFFTRGKQVKGVDVYDERLLVYR 1430

RESULT 5
AAU98042
ID AAU98042 standard; protein; 1430 AA.

XX AC AAU98042;

XX DT 27-AUG-2002 (first entry)

XX S. mutans glucosyltransferase GTFD mutant T589E.

DE DE Glucosyltransferase; GTFD; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture; mutant; mutain.

OS Streptococcus mutans.

OS Synthetic.

XX FH Key Location/Qualifiers

FT Misc-difference 589 /note= "wild-type Thr substituted by Glu"

XX US2002031826-A1.

XX 14-MAR-2002.

XX PF 19-DEC-2000; 2000US-00740274.

XX 07-JUN-1995; 95US-00478704.

PR 07-JUN-1995; 95US-00482711.

PR 07-JUN-1995; 95US-00485243.

PR 16-JAN-1998; 98US-00007999.
PR 16-JAN-1998; 98US-00008172.
PR 20-JAN-1998; 98US-00009620.
PR 11-DEC-1998; 98US-00210361.
XX (NICH/) NICHOLS S E.
XX Nichols SE;
XX WPI; 2002-414332/44.
XX Glucosyltransferase B or D protein useful for producing a glucan useful
PT as substitutes for and additions to modified starch and latexes in paper
PT manufacture, comprises mutations in specific positions.
XX Claim 36; Page: 44pp; English.
XX The invention an isolated protein comprising a glucosyltransferase (GTF)
CC B polypeptide having changes at position from 1448V, D457N, D567T,
CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/D457N, D567T/
CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
CC complementary polynucleotide, a ribonucleic acid sequence encoding the
CC GTF mutant, an expression cassette comprising the polynucleotide operably
CC linked to a promoter, a vector comprising the expression cassette, host
CC cell introduced with the vector, a transgenic plant comprising the
CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
CC coating composition comprising a glucan produced in a plant transformed
CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
CC thermoplastic molecule or their combinations or glucan and starch where
CC the glucan is produced in the amyloplast and/or vacuole or a maize line
CC deficient in starch biosynthesis, transformed with a gene encoding a
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
CC comprising the glucan (paper sizing/coating agent). The vector is useful
CC for producing a glucan in a plant. The method comprises transforming a
CC plant cell with the vector, growing the plant cell under plant growing
CC conditions to produce a regenerated plant and inducing expression of the
CC polynucleotide for a time sufficient to produce the glucan in the
CC regenerated plant, where the vector contains a transit sequence from
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
CC is produced in the amyloplast of potato or the vacuole of sugar beet.
CC Glucans are useful as substitutes for and additions to modified starch
CC and latexes in paper manufacture. Unlike prior art techniques, which
CC require input materials that produce chemical effluents, paper
CC manufacture utilising the glucan produced by GTF, which utilises
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
CC properties and impart gloss to the paper during coating step. The present
CC sequence represents a GTFD mutant of the invention. Note: The present
CC sequence is not shown in the specification but was created by the indexer
CC using the GTFD sequence appearing as AAU98029 and the information in
CC claim 36
XX Sequence 1430 AA;

QY Query Match 99.9%; Score 7456; DB 5; Length 1430;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1429; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 METKRRYKQHKVKKHWTVAVASGLITLTGTTTLGSSVSAETEQTSQSKVVTQKSEDDKAA 60
Db 1 METKRRYKQHKVKKHWTVAVASGLITLTGTTTLGSSVSAETEQTSQSKVVTQKSEDDKAA 60
QY 61 SESSOTDAPKTKQAOTEQTQASQANVADTSTSIKTPSQNITTTQANSDDKTVTNTKSE 120
Db 61 SESSOTDAPKTKQAOTEQTQASQANVADTSTSIKTPSQNITTTQANSDDKTVTNTKSE 120
QY 121 EAQTSSEERTKQSEEAQTASSQALYQAKAELTKQRTAAQENKPNFVDLAAIPNVKQIDGK 180
Db 121 EAQTSSEERTKQSEEAQTASSQALYQAKAELTKQRTAAQENKPNFVDLAAIPNVKQIDGK 180

QY 181 YYYIGSDGPKQKQNFALTVNNKVLVYFDKNTGALTDTSTQYQFKQGLTKLNNDYTPHNQIVNF 240
DB 181 YYYIGSDGPKQKQNFALTVNNKVLVYFDKNTGALTDTSTQYQFKQGLTKLNNDYTPHNQIVNF 240
QY 241 ENTSLKETIDNYTADSWYRKPDLKNGKWTATSESDLRPLLSWMPDKOTQAYLNYMN 300
DB 241 ENTSLKETIDNYTADSWYRKPDLKNGKWTATSESDLRPLLSWMPDKOTQAYLNYMN 300
QY 301 QOGLGTGENYTDSSQESLNLAAQTQVQKLETKISQTOQTQWLRDIINSFVKTPQPNWSQ 360
DB 301 QOGLGTGENYTDSSQESLNLAAQTQVQKLETKISQTOQTQWLRDIINSFVKTPQPNWSQ 360
QY 361 TESDTSAGEKDHLOGGALLYSNDSKTAAYANSYDRLNRPSTSGPKPKYFEDNSSGGYDF 420
DB 361 TESDTSAGEKDHLOGGALLYSNDSKTAAYANSYDRLNRPSTSGPKPKYFEDNSSGGYDF 420
QY 421 LLANDIDNSNPVQVQASQLNLWHLVLMYGSIVANDPEANPDGVRVDVANDVNADLLQIASD 480
DB 421 LLANDIDNSNPVQVQASQLNLWHLVLMYGSIVANDPEANPDGVRVDVANDVNADLLQIASD 480
QY 481 YLKAHYGVDSKSEKNAIHLNLSLEAWSNDPQYNKDTKGAQLPIDNKLRLSLLYALTRPLE 540
DB 481 YLKAHYGVDSKSEKNAIHLNLSLEAWSNDPQYNKDTKGAQLPIDNKLRLSLLYALTRPLE 540
QY 541 KQASNKNEIRSGLEPVITNSLNRSAEGKNSERMANYIFIRAHDSVQTVIAKIIKAQIN 600
DB 541 KQASNKNEIRSGLEPVITNSLNRSAEGKNSERMANYIFIRAHDSVQTVIAKIIKAQIN 600
QY 601 PKTDGTLFTLDELKQAFKLYNEDMRQAKKKTQSNIPATAYALMSLNKDSITRLYYGDMYS 660
DB 601 PKTDGTLFTLDELKQAFKLYNEDMRQAKKKTQSNIPATAYALMSLNKDSITRLYYGDMYS 660
QY 661 DDQGYMATKSPYYDAITLLKARIKYAAGQDMKITTYVEGDKSHMDWDYTGVLTSVRYGT 720
DB 661 DDQGYMATKSPYYDAITLLKARIKYAAGQDMKITTYVEGDKSHMDWDYTGVLTSVRYGT 720
QY 721 GANEATDQGESEATKQGMNAVITSNPNLSKLNQNDKVIIVNNGAAHKNQYRPLLLTTKQGL 780
DB 721 GANEATDQGESEATKQGMNAVITSNPNLSKLNQNDKVIIVNNGAAHKNQYRPLLLTTKQGL 780
QY 781 TSYTSDAAKSLYRKTNDKGLVFDASDIQGYLNPQVSGYLAVWVPVGASDNQDVRVAAS 840
DB 781 TSYTSDAAKSLYRKTNDKGLVFDASDIQGYLNPQVSGYLAVWVPVGASDNQDVRVAAS 840
QY 841 NKANATQGVYESSALDSQLIYEGFSNFODFTVKSDYTNKKIAQNVQLFKSWGVTSPFM 900
DB 841 NKANATQGVYESSALDSQLIYEGFSNFODFTVKSDYTNKKIAQNVQLFKSWGVTSPFM 900
QY 901 APQYVSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSSQDMINAVKALHKSIGQVIAD 960
DB 901 APQYVSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSSQDMINAVKALHKSIGQVIAD 960
QY 961 WVPDQIYNLPKGEVTVATRVNDYGEYKDSSEIKNTLYAANTKSNKDYQAKYGGAPLSL 1020
DB 961 WVPDQIYNLPKGEVTVATRVNDYGEYKDSSEIKNTLYAANTKSNKDYQAKYGGAPLSL 1020
QY 1021 AAKYPSIFNRTQISNGKKIDPSEKITAMKAKYFNGTNIILGRGVYVLDKNASDKYFELKG 1080
DB 1021 AAKYPSIFNRTQISNGKKIDPSEKITAMKAKYFNGTNIILGRGVYVLDKNASDKYFELKG 1080
QY 1081 NOTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFPVQDAKGNWYFDDNNGHWYGL 1140
DB 1081 NOTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFPVQDAKGNWYFDDNNGHWYGL 1140
QY 1141 QQLNGEVQYFLSNGVQLRESFLENADGSKNYFPHGLNRYNSGYYSFDDNSKWRYPDASGV 1200
DB 1141 QQLNGEVQYFLSNGVQLRESFLENADGSKNYFPHGLNRYNSGYYSFDDNSKWRYPDASGV 1200
QY 1201 MAVGLKTINGNTQYFDQGYQVKGAWITGSDGKKYFDDGSGNMAVNFANDKNGDWYLL 1260
DB 1201 MAVGLKTINGNTQYFDQGYQVKGAWITGSDGKKYFDDGSGNMAVNFANDKNGDWYLL 1260

QY 1261 NSDGIALVGQTINGKTYTFEGQDGKQIKGKIITDNGKLKYLFLANSSELARNIFATDSQNN 1320
DB 1261 NSDGIALVGQTINGKTYTFEGQDGKQIKGKIITDNGKLKYLFLANSSELARNIFATDSQNN 1320
QY 1321 WYFSGDGVAVTGSQTIAGKLYFASDGKQVKSFTYNGKLVHYHADSGLQVNRFEAD 1380
DB 1321 WYFSGDGVAVTGSQTIAGKLYFASDGKQVKSFTYNGKLVHYHADSGLQVNRFEAD 1380
QY 1381 KDCNWTYLLDSNGEALTGSQRINDQRFVFFTRREGKQVKGVDVAYDERRLLVYR 1430
DB 1381 KDCNWTYLLDSNGEALTGSQRINDQRFVFFTRREGKQVKGVDVAYDERRLLVYR 1430

RESULT 6
AAU98041
ID AAU98041 standard; protein; 1430 AA.
XX
AC AAU98041;
XX
DT 27-AUG-2002 (first entry)
XX
S. mutans glucosyltransferase GTFD mutant T589D.
XX
Glucosyltransferase; GTFD; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture; mutant; mutein.
XX
OS Streptococcus mutans.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 589 /note= "Wild-type Thr substituted by Asp"
XX
US2002031826-A1.
XX
14-MAR-2002.
XX
19-DEC-2000; 2000US-00740274.
XX
07-JUN-1995; 95US-00478704.
PR 07-JUN-1995; 95US-00482711.
PR 07-JUN-1995; 95US-00485243.
PR 16-JAN-1998; 98US-00007999.
PR 16-JAN-1998; 98US-00008172.
PR 20-JAN-1998; 98US-00009620.
PR 11-DEC-1998; 98US-00210361.
XX
(NICH/) NICHOLS S E.
XX
Nichols SE;
XX
WPI; 2002-414332/44.
XX
Glucosyltransferase B or D protein useful for producing a glucan useful
PT as substitutes for and additions to modified starch and latexes in paper
PT manufacture, comprises mutations in specific positions.
XX
Claim 36; Page; 44pp; English.
XX
The invention an isolated protein comprising a glucosyltransferase (GTF)
CC B polypeptide having changes at position from 1448V, D457N, D567T,
CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,
CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
CC complementary polynucleotide, a ribonucleic acid sequence encoding the
CC GTF mutant, an expression cassette comprising the polynucleotide operably
CC linked to a promoter, a vector comprising the expression cassette, host
CC cell introduced with the vector, a transgenic plant comprising the
CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
CC coating composition comprising a glucan produced in a plant transformed

CC with a gene encoding the mutant GFP, wild type or, starch, a latex,
 CC thermoplastic molecule or their combinations or glucan and starch where
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line
 CC deficient in starch biosynthesis, transformed with a gene encoding a
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
 CC comprising the glucan (paper sizing/coating agent). The vector is useful
 CC for producing a glucan in a plant. The method comprises transforming a
 CC plant cell with the vector, growing the plant cell under plant growing
 CC conditions to produce a regenerated plant and inducing expression of the
 CC polynucleotide for a time sufficient to produce the glucan in the
 CC regenerated plant, where the vector contains a transit sequence from
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.
 CC Glucans are useful as substitutes for and additions to modified starch
 CC and latexes in paper manufacture. Unlike prior art techniques, which
 CC require input materials that produce chemical effluents, paper
 CC manufacture utilising the glucan produced by GFP, which utilises
 CC biologically produced input materials, is more cost-effective and
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
 CC properties and impart gloss to the paper during coating step. The present
 CC sequence represents a GFPD mutant of the invention. Note: The present
 CC sequence is not shown in the specification but was created by the indexer
 CC using the GFPD sequence appearing as AAU98029 and the information in
 CC claim 36
 XX
 SQ Sequence 1430 AA;

Query Match 99.9%; Score 7456; DB 5; Length 1430;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1429; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 METKRRYKMHKVKHWHVAVASGLITLTGTTTILGSSVSAETEQSDTKVVTQKSEDDKAA 60
 Db 1 METKRRYKMHKVKHWHVAVASGLITLTGTTTILGSSVSAETEQSDTKVVTQKSEDDKAA 60

QY 61 SESSQTDAPKTKQAOTEQTAQOSQANVADTSTSIKETPSONITTOANSDDKTVNTKSE 120
 Db 61 SESSQTDAPKTKQAOTEQTAQOSQANVADTSTSIKETPSONITTOANSDDKTVNTKSE 120

QY 121 EAQTSHERTKQSEEAQTASSQALTOAKAELTKQRTAAQENKPNVDLAAIPNVKQIDGK 180
 Db 121 EAQTSHERTKQSEEAQTASSQALTOAKAELTKQRTAAQENKPNVDLAAIPNVKQIDGK 180

QY 181 YYYIGSDGQPKNFALTNNKVLFPDKNTGALTDTTSQYQFKGLTKLNDYTPHNQIVNF 240
 Db 181 YYYIGSDGQPKNFALTNNKVLFPDKNTGALTDTTSQYQFKGLTKLNDYTPHNQIVNF 240

QY 241 ENTSLETTIDNYVTADSWRPKDILKNGKTTWTASSSDLRPLLSWMPDKQTOIAYLNYNN 300
 Db 241 ENTSLETTIDNYVTADSWRPKDILKNGKTTWTASSSDLRPLLSWMPDKQTOIAYLNYNN 300

QY 301 QQGLGTGENYTADSSQESLNLAQTQVVKIETKISQTOQTOMLRDILNSFVKTPQWNNSQ 360
 Db 301 QQGLGTGENYTADSSQESLNLAQTQVVKIETKISQTOQTOMLRDILNSFVKTPQWNNSQ 360

QY 361 TESDTSAGEKHLQGGALLYSNSDKTAYANSYRLNRTPTSTQTKPKYFEDNSSGGYDF 420
 Db 361 TESDTSAGEKHLQGGALLYSNSDKTAYANSYRLNRTPTSTQTKPKYFEDNSSGGYDF 420

QY 421 LLANDIDNSNPVQAEQNLWHLNMYNYSIVANDPEANPDGVRVDAVDNWNADLLQIASD 480
 Db 421 LLANDIDNSNPVQAEQNLWHLNMYNYSIVANDPEANPDGVRVDAVDNWNADLLQIASD 480

QY 481 YLKAHYGVDDKSEKNAIHLNLSILEANSNDPQNKDTKGAQLPDINKLRSLLYALTRPLE 540
 Db 481 YLKAHYGVDDKSEKNAIHLNLSILEANSNDPQNKDTKGAQLPDINKLRSLLYALTRPLE 540

QY 541 KDAASKNEIRSGLEPVITNSLNNRNSAEGKNSRMANYIPFIRAHDSVQTVIAKIIKAQIN 600
 Db 541 KDAASKNEIRSGLEPVITNSLNNRNSAEGKNSRMANYIPFIRAHDSVQTVIAKIIKAQIN 600

QY 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKKYTQSNIPATAYALMLSNKDSITRLYYGDMYS 660

Db 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKKYTQSNIPATAYALMLSNKDSITRLYYGDMYS 660
 QY 661 DDGQYMATKSPYDAIDTLLKARIKYAAGGQDMKITTYVEGDKSHMDWDYTGVLTSVRYGT 720
 Db 661 DDGQYMATKSPYDAIDTLLKARIKYAAGGQDMKITTYVEGDKSHMDWDYTGVLTSVRYGT 720
 QY 721 GANEATDQGSSEATKTQGMNAVITSNPSSLKLNQNDKVIIVNMGAAHKRQYRPLLLTTKQGL 780
 Db 721 GANEATDQGSSEATKTQGMNAVITSNPSSLKLNQNDKVIIVNMGAAHKRQYRPLLLTTKQGL 780
 QY 781 TSYSDAAAKSLYKRTNDKGLVFPDASDIQGVLPQVSGYLAVWPVPGASDNQDVRVAAS 840
 Db 781 TSYSDAAAKSLYKRTNDKGLVFPDASDIQGVLPQVSGYLAVWPVPGASDNQDVRVAAS 840
 QY 841 NKANATGOVYESSSALDSQLIYEGFSNFQDFVTKDSYTNKKIAQNVLFKSWGVTSPFM 900
 Db 841 NKANATGOVYESSSALDSQLIYEGFSNFQDFVTKDSYTNKKIAQNVLFKSWGVTSPFM 900
 QY 901 APQYVSSSEGGFSLDSIIQNGYAFEDRYDLAMSKNNKYSQQDMINAVKALHKSIGQVIAD 960
 Db 901 APQYVSSSEGGFSLDSIIQNGYAFEDRYDLAMSKNNKYSQQDMINAVKALHKSIGQVIAD 960
 QY 961 WVPDQIYNLPGKEVVVTATRVNDYGEYRKDSEIKNTLYAANTKSNKGDYQAKYGGAFSLSEL 1020
 Db 961 WVPDQIYNLPGKEVVVTATRVNDYGEYRKDSEIKNTLYAANTKSNKGDYQAKYGGAFSLSEL 1020
 QY 1021 AAKYPSIFNRTQISNGKKIDPSEKITAWKAKYFNGTNIILGRGVYLVKDNASDKYFELKG 1080
 Db 1021 AAKYPSIFNRTQISNGKKIDPSEKITAWKAKYFNGTNIILGRGVYLVKDNASDKYFELKG 1080
 QY 1081 NOTYLPKQMTNKEASTGFVNDGNGMTFYSTGYSQAQNSFVQDAKGNWYTFDNNGHMVYGL 1140
 Db 1081 NOTYLPKQMTNKEASTGFVNDGNGMTFYSTGYSQAQNSFVQDAKGNWYTFDNNGHMVYGL 1140
 QY 1141 QOLNGEVQYFLSNGVQLRESFLENADGSKNYFGHLGNYSNGYYSFNDNSKWRYFDASGV 1200
 Db 1141 QOLNGEVQYFLSNGVQLRESFLENADGSKNYFGHLGNYSNGYYSFNDNSKWRYFDASGV 1200
 QY 1201 MAVGLKTINGNTQYFPDQGYQVKGAWITGSDGKKRYFDDGSGNMAVNRFANDKNGDWYYL 1260
 Db 1201 MAVGLKTINGNTQYFPDQGYQVKGAWITGSDGKKRYFDDGSGNMAVNRFANDKNGDWYYL 1260
 QY 1261 NSDGIALVGVQTINGKTYTFGQDGHQIKGIITDNGKLYFLANSSELARNIFATDSQNN 1320
 Db 1261 NSDGIALVGVQTINGKTYTFGQDGHQIKGIITDNGKLYFLANSSELARNIFATDSQNN 1320
 QY 1321 WYFSGDGVAVTGSOTIAGKKLYFASDQGVKQKSFVYNGKVHYHADSGELQVNRFEAD 1380
 Db 1321 WYFSGDGVAVTGSOTIAGKKLYFASDQGVKQKSFVYNGKVHYHADSGELQVNRFEAD 1380
 QY 1381 KQGNWYILDSNGEALTGSRINDQRVFFTFREGQVKGVDVAYDERLLVYR 1430
 Db 1381 KQGNWYILDSNGEALTGSRINDQRVFFTFREGQVKGVDVAYDERLLVYR 1430

RESULT 7
 AAU98044
 ID AAU98044 standard; protein; 1430 AA.
 XX
 AC AAU98044;
 XX
 DT 27-AUG-2002 (first entry)
 XX
 DE S. mutans glucosyltransferase GTFD mutant N471D/T589D.
 XX
 KW Glucosyltransferase; GTFD; transgenic plant; paper sizing;
 KW coating composition; glucan; starch; latex; thermoplastic molecule;
 XX amyloplast; vacuole; paper manufacture; mutant; mutain.
 OS Streptococcus mutans.
 OS Synthetic.

PH	Key	Location/Qualifiers
FT	Misc-difference 471	
FT	/note= "Wild-type Asn substituted by Asp"	
FT	Misc-difference 589	
FT	/note= "Wild-type Thr substituted by Asp"	
XX		
XX	US2002031826-A1.	
XX		
XX	14-MAR-2002.	
XX		
XX	19-DEC-2000; 2000US-00740274.	
XX		
PR	07-JUN-1995; 95US-00478704.	
PR	07-JUN-1995; 95US-00482711.	
PR	07-JUN-1995; 95US-00485243.	
PR	16-JAN-1998; 98US-00007999.	
PR	16-JAN-1998; 98US-00008172.	
PR	20-JAN-1998; 98US-00009620.	
PR	11-DEC-1998; 98US-00210361.	
XX		
FA	(NICH/) NICHOLS S E.	
XX		
PI	Nichols SE;	
XX		
XX	WPT; 2002-414332/44.	
DR		
XX		
FT	Glucosyltransferase B or D protein useful for producing a glucan useful	
FT	as substitutes for and additions to modified starch and latexes in paper	
PT	manufacture, comprises mutations in specific positions.	
XX		
XX	Claim 36; Page; 44pp; English.	
XX		
CC	The invention an isolated protein comprising a glucosyltransferase (GTF)	
CC	B polypeptide having changes at position from I448V, D457N, D567T,	
CC	K1014T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,	
CC	I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a	
CC	GTF D polypeptide having changes at positions from T599D, T599E, N471D,	
CC	N471D/T599D, and N471D/T599E. Also included are a glucan produced by the	
CC	GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its	
CC	complementary polynucleotide, a ribonucleic acid sequence encoding the	
CC	GTF mutant, an expression cassette comprising the polynucleotide operably	
CC	linked to a promoter, a vector comprising the expression cassette, host	
CC	cell introduced with the vector, a transgenic plant comprising the	
CC	vector, a seed or tuber from the transgenic plant, a paper sizing and/or	
CC	coating composition comprising a glucan produced in a plant transformed	
CC	with a gene encoding the mutant GTF, wild type or, starch, a latex,	
CC	thermoplastic molecule or their combinations or glucan and starch where	
CC	the glucan is produced in the amyloplast and/or vacuole or a maize line	
CC	deficient in starch biosynthesis, transformed with a gene encoding a	
CC	glucosyltransferase B or D enzyme, wild-type or mutant and a paper	
CC	comprising the glucan (paper sizing/coating agent). The vector is useful	
CC	for producing a glucan in a plant. The method comprises transforming a	
CC	plant cell with the vector, growing the plant cell under plant growing	
CC	conditions to produce a regenerated plant and inducing expression of the	
CC	polynucleotide for a time sufficient to produce the glucan in the	
CC	regenerated plant, where the vector contains a transit sequence from	
CC	ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and	
CC	chlorophyll AB binding protein to produce a transgenic plant, and glucan	
CC	is produced in the amyloplast of potato or the vacuole of sugar beet.	
CC	Glucans are useful as substitutes for and additions to modified starch	
CC	and latexes in paper manufacture. Unlike prior art techniques, which	
CC	require input materials that produce chemical effluents, paper	
CC	manufacture utilising the glucan produced by GTF, which utilises	
CC	biologically produced input materials, is more cost-effective and	
CC	environmentally friendly. Moreover, glucans also exhibit thermoplastic	
CC	properties and impart gloss to the paper during coating step. The present	
CC	sequence represents a GTFD mutant of the invention. Note: The present	
CC	sequence is not shown in the specification but was created by the indexer	
CC	using the GTFD sequence appearing as AAU98029 and the information in	
CC	claim 36	
XX		
XX	Sequence 1430 AA;	
XX		

Db 1021 AAKYPSIFNRTQISNGKIDPSEKITAWEKAFNGNTILGRGVYGLKDNASDKYFELKG 1080
QY 1081 NQTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYFYFDNNGHMVYGL 1140
Db 1081 NQTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYFYFDNNGHMVYGL 1140
QY 1141 QQLNGEVQYFLSNGVQLRESFLENADGSKNYFGHLGNRYSGYYSFDNDSKWRYFDASGV 1200
Db 1141 QQLNGEVQYFLSNGVQLRESFLENADGSKNYFGHLGNRYSGYYSFDNDSKWRYFDASGV 1200
QY 1201 MAVGLKTINGNTQYFDQDGYQVKGAWITGSDCKRYFDDGSGMAVNRFPANDKNGDWYVL 1260
Db 1201 MAVGLKTINGNTQYFDQDGYQVKGAWITGSDCKRYFDDGSGMAVNRFPANDKNGDWYVL 1260
QY 1261 NSDGIALGVQTINGKTYFYFGDQKQIKGKIITDNGKLYFLANSSELARNIFATDSQNN 1320
Db 1261 NSDGIALGVQTINGKTYFYFGDQKQIKGKIITDNGKLYFLANSSELARNIFATDSQNN 1320
QY 1321 WTYFGSDGVAVNGSQTIAKKLYFASDGKQVKGSGFTYNGKHYHYHADSGELQVNRFEAD 1380
Db 1321 WTYFGSDGVAVNGSQTIAKKLYFASDGKQVKGSGFTYNGKHYHYHADSGELQVNRFEAD 1380
QY 1381 KQGNWYLDNSGEALTGSRINDQVFFTRREGKQVKGDVAYDERLLVYR 1430
Db 1381 KQGNWYLDNSGEALTGSRINDQVFFTRREGKQVKGDVAYDERLLVYR 1430

RESULT 8
AAU98045
ID AAU98045 standard; protein; 1430 AA.
XX AAU98045;
AC AAU98045;
XX
DT 27-AUG-2002 (first entry)
XX
DE S. mutans glucosyltransferase GTFD mutant N471D/T589E.
XX
KW Glucosyltransferase; GTFD; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture; mutant; mutein.
XX
OS Streptococcus mutans.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 471
FT Misc-difference 589 /note= "Wild-type Asn substituted by Asp"
FT Misc-difference 589 /note= "Wild-type Thr substituted by Glu"
XX
PN US2002031826-A1.
XX
PD 14-MAR-2002.
XX
PF 19-DEC-2000; 2000US-00740274.
XX
PR 07-JUN-1995; 95US-00478704.
PR 07-JUN-1995; 95US-00482711.
PR 07-JUN-1995; 95US-00485243.
PR 16-JAN-1998; 98US-00007999.
PR 16-JAN-1998; 98US-00008172.
PR 20-JAN-1998; 98US-00009620.
PR 11-DEC-1998; 98US-00210361.
XX
PA (NICH/) NICHOLS S E.
XX
PI Nichols SE;
XX
XX WPI; 2002-414332/44.
XX
XX Glucosyltransferase B or D protein useful for producing a glucan useful
PT as substitutes for and additions to modified starch and latexes in paper

PT manufacture, comprises mutations in specific positions.
XX Claim 36; Page; 44pp; English.
XX
CC The invention an isolated protein comprising a glucosyltransferase (GTF)
CC B polypeptide having changes at position from 1448V, D457N, D567T,
CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/D1014T,
CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
CC complementary polynucleotide, a ribonucleic acid sequence encoding the
CC GTF mutant, an expression cassette comprising the polynucleotide operably
CC linked to a promoter, a vector comprising the expression cassette, host
CC cell introduced with the vector, a transgenic plant comprising the
CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
CC coating composition comprising a glucan produced in a plant transformed
CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
CC thermoplastic molecule or their combinations or glucan and starch where
CC the glucan is produced in the amyloplast and/or vacuole or a maize line
CC deficient in starch biosynthesis, transformed with a gene encoding a
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
CC comprising the glucan (paper sizing/coating agent). The vector is useful
CC for producing a glucan in a plant. The method comprises transforming a
CC plant cell with the vector, growing the plant cell under plant growing
CC conditions to produce a regenerated plant and inducing expression of the
CC polynucleotide for a time sufficient to produce the glucan in the
CC regenerated plant, where the vector contains a transit sequence from
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
CC is produced in the amyloplast of potato or the vacuole of sugar beet.
CC Glucans are useful as substitutes for and additions to modified starch
CC and latexes in paper manufacture. Unlike prior art techniques, which
CC require input materials that produce chemical effluents, paper
CC manufacture utilising the glucan produced by GTF, which utilises
CC biologically produced input materials, is more cost-effective and
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
CC properties and impart gloss to the paper during coating step. The present
CC sequence represents a GTFD mutant of the invention. Note: The present
CC sequence is not shown in the specification but was created by the indexer
CC using the GTFD sequence appearing as AAU98029 and the information in
CC claim 36
XX
SQ Sequence 1430 AA;
Query Match 99.9%; Score 7451; DB 5; Length 1430;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1428; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 METKRYKHKVKKHVVTVAVASGLITLTGTTLGSSVSAETEQOTSDDKVVTKSEDDKAA 60
Db 1 METKRYKHKVKKHVVTVAVASGLITLTGTTLGSSVSAETEQOTSDDKVVTKSEDDKAA 60
QY 61 SESSOTDAPKTKQAOTEQTAQSQANVADTSTSIKTEPSONITTTQANSDDKVTNTKSE 120
Db 61 SESSOTDAPKTKQAOTEQTAQSQANVADTSTSIKTEPSONITTTQANSDDKVTNTKSE 120
QY 121 EAQTSSEERTKQSEEAQTASSQALTOAKAELTKQRTAAQENKNPVDLAAIPNVKIDGK 180
Db 121 EAQTSSEERTKQSEEAQTASSQALTOAKAELTKQRTAAQENKNPVDLAAIPNVKIDGK 180
QY 181 YYYIGSDGQPKKNFALTNNKVLFPDKNTGALTDTFSQYQFQKGLTKLNDYTPHNOIVNF 240
Db 181 YYYIGSDGQPKKNFALTNNKVLFPDKNTGALTDTFSQYQFQKGLTKLNDYTPHNOIVNF 240
QY 241 ENTSLTETIDNVYTADSWYRPKDILKNGKTTWATSSSESDLRPLMSWWPDKQTOIAYLYNN 300
Db 241 ENTSLTETIDNVYTADSWYRPKDILKNGKTTWATSSSESDLRPLMSWWPDKQTOIAYLYNN 300
QY 301 QOGLGTGENYTADSSQESLNLAATQVQVKIEFKISQTOQTQWLRLDIINSFVTKQPNWNSQ 360
Db 301 QOGLGTGENYTADSSQESLNLAATQVQVKIEFKISQTOQTQWLRLDIINSFVTKQPNWNSQ 360

QY 361 TESDTAGEKHQOGGALLYSNDKTYAYNSDYRLNRTPTSGTKPKYPEDNSSGGYDF 420
DB |||||
QY 361 TESDTAGEKHQOGGALLYSNDKTYAYNSDYRLNRTPTSGTKPKYPEDNSSGGYDF 420
DB |||||
QY 421 LLANDINSPVQAEOQLNHLVLMYVGSIVANDPEANFDGVRDADVNDVADLLQIASD 480
DB |||||
QY 421 LLANDINSPVQAEOQLNHLVLMYVGSIVANDPEANFDGVRDADVNDVADLLQIASD 480
DB |||||
QY 481 YLKAHYGVDSKSEKNAIHLSILEAWSNDPQYNKDTKGAQLPIDNKLRLSLLYALTRPLE 540
DB |||||
QY 481 YLKAHYGVDSKSEKNAIHLSILEAWSNDPQYNKDTKGAQLPIDNKLRLSLLYALTRPLE 540
DB |||||
QY 541 KDSNKNRISGLPEVITNSLNRSAGKNSRMANYIFIRAHDSVQTVIAKIIKAQIN 600
DB |||||
QY 541 KDSNKNRISGLPEVITNSLNRSAGKNSRMANYIFIRAHDSVQTVIAKIIKAQIN 600
DB |||||
QY 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYTOSNIPTAYALMLSKNDSITRLYYGDMYS 660
DB |||||
QY 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYTOSNIPTAYALMLSKNDSITRLYYGDMYS 660
DB |||||
QY 661 DDQYMATKSPYYDAIDTLLKARIKYAAGQDMKITYVEGDKSHMDWDYTGVLTSVRYGT 720
DB |||||
QY 661 DDQYMATKSPYYDAIDTLLKARIKYAAGQDMKITYVEGDKSHMDWDYTGVLTSVRYGT 720
DB |||||
QY 721 GANEATDQSGEATKTOGMAVITSNPDLKNDKVIYVNGAAHKNQYRPLLLTKDGL 780
DB |||||
QY 721 GANEATDQSGEATKTOGMAVITSNPDLKNDKVIYVNGAAHKNQYRPLLLTKDGL 780
DB |||||
QY 781 TSYTSDAKSLYRKNTDNGELVFDASDIQGYLNPQVSGYLAVVPGASDNQDVRVAAS 840
DB |||||
QY 781 TSYTSDAKSLYRKNTDNGELVFDASDIQGYLNPQVSGYLAVVPGASDNQDVRVAAS 840
DB |||||
QY 841 NKANATQGVYESSALDSQLIYEGFSNFODFVTKDSYTNKKAQNVQLFKSWGVTSPRM 900
DB |||||
QY 841 NKANATQGVYESSALDSQLIYEGFSNFODFVTKDSYTNKKAQNVQLFKSWGVTSPRM 900
DB |||||
QY 901 APQYVSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSQQDMINAVKALHKSIGQVIAD 960
DB |||||
QY 901 APQYVSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSQQDMINAVKALHKSIGQVIAD 960
DB |||||
QY 961 WYPDQIYNLPGEVVTATRVNDYGEYRKDSEIKNTLYAANTYNSKNDYQAKYGGAPLSL 1020
DB |||||
QY 961 WYPDQIYNLPGEVVTATRVNDYGEYRKDSEIKNTLYAANTYNSKNDYQAKYGGAPLSL 1020
DB |||||
QY 1021 AAKYPSIFNRTQISNGKIDPSEKITAWKAKYFNGTILGRGVYVLDKNASDKYFELKG 1080
DB |||||
QY 1021 AAKYPSIFNRTQISNGKIDPSEKITAWKAKYFNGTILGRGVYVLDKNASDKYFELKG 1080
DB |||||
QY 1081 NQTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYFFDNNGHMYYGL 1140
DB |||||
QY 1081 NQTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYFFDNNGHMYYGL 1140
DB |||||
QY 1141 QQLNGEVQYFLSNGVQLRESFLENADGSKNYFGLGNRYSGNYGYPFNDNSKRYFDASGV 1200
DB |||||
QY 1141 QQLNGEVQYFLSNGVQLRESFLENADGSKNYFGLGNRYSGNYGYPFNDNSKRYFDASGV 1200
DB |||||
QY 1201 MAVGLKTINGNTQYFDQDQYQVKGAWITGSDGKRYFDDGSGMAVNRFANDKNGDWWYL 1260
DB |||||
QY 1201 MAVGLKTINGNTQYFDQDQYQVKGAWITGSDGKRYFDDGSGMAVNRFANDKNGDWWYL 1260
DB |||||
QY 1261 NSDGIALGVQTINGKTYFFGDGQKQIKGKIITDNGKLYFLANSSELARNIFATDSQNN 1320
DB |||||
QY 1261 NSDGIALGVQTINGKTYFFGDGQKQIKGKIITDNGKLYFLANSSELARNIFATDSQNN 1320
DB |||||
QY 1321 WYFSGDGVAVTGSQTIAGKLYFPASDGHQVKGFSFVYTKNGKHYHYTHASGELQVNRFEAD 1380
DB |||||
QY 1321 WYFSGDGVAVTGSQTIAGKLYFPASDGHQVKGFSFVYTKNGKHYHYTHASGELQVNRFEAD 1380
DB |||||
QY 1381 KDGNNYLLDSNGEALTGSRINDQRVFFTRREGQVKGVDVAYDERLLVYR 1430
DB |||||
QY 1381 KDGNNYLLDSNGEALTGSRINDQRVFFTRREGQVKGVDVAYDERLLVYR 1430
DB |||||

RESULT 9

AAU79284

ID AAU79284 standard; protein; 1476 AA.

XX AC AAU79284;

XX XX 13-AUG-2002 (first entry)

XX XX Streptococcus mutans monoclonal antibody-related protein #1.

XX XX Antibody; dental caries; water insoluble glucan synthetase; anti-caries;

XX XX glucosyl transferase-B; immunotherapy.

XX XX Streptococcus mutans.

XX XX JP20021114709-A.

XX PD 16-APR-2002.

XX PF 04-OCT-2000; 2000JP-00304889.

XX PR 04-OCT-2000; 2000JP-00304889.

XX XX (UYNI-) UNIV NIPPON.

XX XX WPI; 2002-448885/48.

XX XX Anti-caries agent composed of a monoclonal antibody against an inhibitory

XX XX enzyme against water insoluble glucan synthetase of glucosyl transferase-

XX XX B (GTF-B) of Streptococcus mutans.

XX PS Claim 3; Page 13-16; 28pp; Japanese.

XX XX The invention relates to a monoclonal antibody against dental caries and

XX XX an anti-caries agent composed of a monoclonal antibody produced by

XX XX Streptococcus mutans, particularly mouse-hybridoma MHP126 (FERM P-17566)

XX XX or mouse-hybridoma MHP136 (FERM P-17567), against an enzyme having

XX XX inhibitive activity against water insoluble glucan synthetase of glucosyl

XX XX transferase-B. The monoclonal antibody specifically inhibits water

XX XX insoluble glucan synthetase of Streptococcus mutans produced glucosyl

XX XX transferase-B and is used in the immunotherapy of dental caries. This

XX XX sequence represents a Streptococcus mutans monoclonal antibody-related

XX XX protein

XX SQ Sequence 1476 AA;

Query Match 51.8%; Score 3869; DB 5; Length 1476;

Best Local Similarity 53.3%; Pred. No. 5.4e-192;

Matches 767; Conservative 229; Mismatches 394; Indels 48; Gaps 20;

QY 1 METKERYKMHKVKHMTVAVASGLITLTGLSSVSAETEQOTSDKVVTKSEDDKAA 60

DB 1 MDKRYKLRKVKRGTWSVASVMTL-TTLSSGLVADSNESKS-----QISNDSNTS 54

QY 61 SSSQTDAPKTKQAQTEQTAQOSQAN-VADTSTSTITKETPSONITTOANSDDKTVNTKS 119

DB 55 VTANEESNVITEATSKQEAASSQTHVTVSSSTSVVNPKEVSNPTVGTETASNGEK 114

QY 120 BEAQTSEBRTKQSEBAQTASSQALTOAKAELTKORQTAAGNKNPVDLAALIPNVKQIDG 179

DB 115 LQNQTTV-DKTSEAAANNISKQT-TEADTDVIDDSNAA-----NLQILEKLPNVKEIDG 167

QY 180 KYIYIGSDGQPKNFALTNNKLVYFDKNTGALTDTTS-QYQFKQGLTKLNNDTYTHNQIV 238

DB 168 KYIYDNNKGRVNTFTLIADGKILHFE-TGAYTDTSDTVNKDITVTRSNLYKKYNQY 226

QY 239 NPENTSLETIONVTADSWYRPKDILKNGKWTASSESDRLPLMSWPPDKOTQIAYLNY 298

DB 227 DRSAQSPFHVDRHLYTAESWYRPKYLKKGKWTQSTEKDFRPLLMWPPDQSTQYVNY 286

QY 299 MNQOGLGTGENYTADSSQESLNAAQTVQVKIETKISQTOQTQWLDRDIINSFVKTPQNN 358

DB 287 MNAQ-LGINKTYDDTSNQLNIAAATIQAKIEAKITTLKNTDLRQITSAFVKTSQAWN 345

QY 359 SOTSEDSAGEKHLOGGALLYSNDK-TAYANSYRLNRTPTQTK--PKYPEDNSS 415
Db 346 SDESEKPF-----DHLQNGAVLYDNEGKLPYANSNYRLNRTPTQTKKDDPRYADNTI 401
QY 416 GGYDFLLANDIONSPPVQAEOLNWLHLYLMNYGSIIVANDPEANFDGVRVDVADNVDNADLL 475
Db 402 GGYEFLLANDVNSNPVQAEOLNWLHFLMNFNIYANDPDANFDSIRVDVADNVDADLL 461
QY 476 QIASDYLKAHYGVDSKNAIHNLSILEAWSNDPQYNKDTKGAQLPIDNKRLRLSLLYAL 535
Db 462 QIAGDYLLKAAKGTHKNDKAANDHLSILEAWSNDPPLYHDDGDNDINMDNKLRLSLLFSL 521
QY 536 TRPLEKDAANKNEIRSCLEPVTNSLNRSAGKNSERMANIYIFRAHDSVQTVIAKII 595
Db 522 AKPLNQ-----RSGMNPILTNSLVNRTDDNAETAAPVSYSFIRAHDSVQDLIRDI 573
QY 596 KQAINPKTDGLFTLDELKQAFKYNEDMRQAKKYTQSNIPAYVALMSNKDSITRLY 655
Db 574 KAEINPNVVGYSFTWEEIKKAFEYINKDLJATEKKITHYNTALSTALLTNKSSVPRVY 633
QY 656 GMYSDGQYMATKSPYDAIDTLKARIKYAAGGQDMKITTYVEGDKSHMDWDYTGVLTS 715
Db 634 GDMFTDDGQYMAKHTINYEAETLLKARIKYVSGGQAMENQOVGNSE-----IITS 684
QY 716 VRYGTGANBATOGESEATYTOGMVITSNPCLKNDKVINMGAAHQEQEYRPLLT 775
Db 685 VRYGKALKATTDGTRTTRTSVGAVIEGNNPRLKASDRVVVNMGAHKNAQYRPLLT 744
QY 776 TKDGLTSYSDAAKSLYKTKNDKGBELVFDASDIQGYLNPQVSGYLAVWVPVPGASDNQDV 835
Db 745 TONGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKYANPQVSGYLGWVPVGAADQDV 804
QY 836 RVAASNKANATQVYBESSALDSQIYEGFSNFQDFVTTKDSYTNKKIAQNVQLFKSGWV 895
Db 805 RVAASTAPSTDGKSVHQAALDSRVMEGFSNFQAFATKKEEYTNVIAKNVDKFAEMGV 864
QY 896 TSFEMAPQVSSDEGDFLSIITONGYAFEDRVDLAMSKNKYGSQODMNAVKAHKSIGI 955
Db 865 TDFEMAPQVSSDGSFLDSVQNGYAFTRDYDLGISEKNKYGTADDLVKAIKALHSGKI 924
QY 956 QVIADVPQIYNLPGKVVVTRVNDYGEYKDSIKNTLYAANTKSGDKDYQAKYGA 1015
Db 925 KYMADVPQMTAFPEKEVVTIRVDKYGTVPAGSQIKNLTIVDGKSGKQQAQYGA 984
QY 1016 FLSELAQYPSIFNRTOISNGKKIDPSEKITAFAKYPNGTILGRGVGYVLKDNASDKY 1075
Db 985 FLEELQAKYPELFARQIISTGVPMDDPSVKIKQWSAKYFNGTILGRGAGYVLKQQAATNY 1044
QY 1076 FELKGNQ--TYLPKQMTNKEASTGVNDGNGMTFTYSTGYQAKNSFVQDQAKGNWYFDNN 1133
Db 1045 FNISDNKEINFLPKTLNLDQSVQGSYDGKGYVYSTGYQAKNTPISEG-DKWYFYDNN 1103
QY 1134 GHMVTGLQNGEVOYFLSNGVQLRESFLENADGSKNYFGLGNRYSGNYGYSFNDNSKWR 1193
Db 1104 GTMVTGAQINGVNYFLSNGVQLRDLILKNEDEGTVAIYNGDORRYENGYOF-MSGWR 1162
QY 1194 YFDASGVMAVGLKTNGNTQYFDQDGYQVKGAWITGSDGKRYFDDGSGNMAVNRFANDK 1253
Db 1163 HFN-NGEMSGLTVIDGQVQYFDEMGYQAKGFVTTADGKIRYFDKQSGNMYNRPIENE 1221
QY 1254 NGDWYLSNGIALGVQVQTINGKYTYFFQDQKQIKGIITD-NGKKUYFLANSBELARNI 1312
Db 1222 EGKWLVLGDEGAVTGSQTINGQHLVFRANGVQVKGFEVTDHHRISYYDNGSGDQIRNR 1281
QY 1313 FATDSNNWYFGSDGAVTGSQTIAGKKLYPASDGKQVKGFSVTYN-GKVHYHYADSGE 1371
Db 1282 FVRNAQGVYFDNNGYAVTGARTINGQHLVFRANGVQVKGFEVTDRIYGRISYYDNGSGD 1341
QY 1372 LQVNRPEADKGNWYLYDNGSEALTGSRINDQVREFTREGQVKGVDVADYERLLVY 1429
Db 1342 QIRNRFPVRNAQGVYFDNNGYAVTGARTINGQHLVFRANGVQVKGFEVTDRIYGRISY 1399

RESULT 10
ADC54806
ID ADC54806 standard; protein; 1499 AA.
XX AC ADC54806;
XX AC ADC54806;
DT 18-DEC-2003 (first entry)
XX Protein Seq ID11 related to L mesenteroides dextran sucrose protein.
XX dextran sucrose; active centre zone; glucan; polysaccharide; dextran;
KW D-glucose; starch; cellulose; glucan manufacture; transduction;
KW enzyme-reaction product.
XX Unidentified.
XX JP2003111590-A.
XX PN 15-APR-2003.
XX PD 03-OCT-2001; 2001JP-00307067.
XX PF 03-OCT-2001; 2001JP-00307067.
XX PR 03-OCT-2001; 2001JP-00307067.
XX PA (DOKU-) DOKURITSU GYOSEI HOJIN SHOKUJIN SOGO KEN.
XX DR WPI; 2003-735670/70.
XX P-PSDB; ADC54814.
XX Novel modified dextran sucrose which exchanges one site of active center
zone of dextran sucrose for active center zone of different types of
dextran sucrose, useful for manufacturing glucan.
XX Example 1; SEQ ID NO 11; 28pp; Japanese.
XX This invention relates to a modified dextran sucrose (DS) exchanging one
site of the active centre zone of a dextran sucrose for the active centre
zone of different types of dextran sucrose. Glucan is a polysaccharide
(for example dextran) which uses D-glucoses, such as a starch and a
cellulose, as a structural unit. The modified enzyme of the invention is
useful in the manufacture of glucan. The selection of the active centre
area of DS which carries out transduction, enables changes in structure
and character of an enzyme-reaction product and their application to
various uses. The present sequence is that of a protein which is related
to the leuconostoc mesenteroides dextran sucrose protein and which was
used during the exemplification of the invention.
SQ Sequence 1499 AA;
Query Match 51.8%; Score 3867; DB 7; Length 1499;
Best Local Similarity 53.4%; Pred. No. 7e-192;
Matches 775; Conservative 202; Mismatches 415; Indels 58; Gaps 19;
QY 17 VTVAVASGLITLGTTLGSSV-----SARTEQOTSQDKVVT--QKSEDDKAA--SSSQTD 67
Db 36 VVAVAVQSNLDATSDSKSITTDKAAATATSTDDKATTTADTTDDKAAATATSTDD 95
QY 68 APKTKQAQTEQTAQOQANVADTSTSTKETPSQN--ITTOANS-DDKVTVN--TKSEBA 122
Db 96 KATTTAATSTDDKATTAATSTDDKATTTADTTDDKATTTAATSTDDKATTTAATSTDDK 155
QY 123 OTSEERTKQSEBAQTA-----SSQALTQAKAELTKQRTAAQENK 163
Db 156 ATTTAATSTDDKAAATTAATSTDDKATTTADTTDDKATTTAATSTDDKATTTAATSTDDK 215
QY 164 NPVDLAAI-----PNVKQIDGKYYIGSDGPKKNFALTVNNKLVDFKNTGALTDTSOYQ 219
Db 216 TADTTAALEASKNLKTIDGKTYIYDDDDQVKKNFATVIDGKLVYDFDKETGALLATNDYQ 275
QY 220 FKQGLTKLNDYTPHNQIVNFENTSETIDNYVTADSWYRPKDILKNGKWTWASSSDLR 279
Db 276 FLEGITSENNTYTEHNASVGTTSDSYTNVDGYLTADSWYRPKDILVNGQWSSSKDDDLR 335


```
QY 280 PLLMSWPKQOTQIAYLYNNQO--OGLGTGENYTTADSSQESLNLAQAOTVQVKIETKISQTQ 338
Db 336 PLLMTWPKDKATQVNLNAMKYLDATEETETVYTSDDSDALNKAQAQVQVKEIKISQEG 395
QY 339 QTCMLRDIINSFVKYTPQNNQSOTESDTSAGEKHLQGGALLYSNDSKTAYANSDYLLNR 398
Db 396 QTCMLKDDISKFDVSSNNNIASESGT----DHQGGALLYNSDKTDPANSDYLLNR 451
QY 399 TPTSQTGKPKYFEDNSGGYDFLLANDIDNSNPVQAEQLNLWLHLYLMTNYSIIVANDPEAN 458
Db 452 TPTNQGTPLTYTDPQTGGYDFLLANDVNSNPVQAEQLNLWYLLNFGSITNNADAN 511
QY 459 FQGVVDVAVNNADLLQIASDYLYKAHYGVDSKNAHLSILEAWSNDPQYNKDTKG 518
Db 512 FDSIRVDVAVNNADLLQIAADYFKAAYGVDSKNAHLSILEAWSNDPQYNKDTKG 571
QY 519 AOLPINDKLRLSLLYALTPLKXDAKNQNEIRSGLEPVTNSLNRSAGKNSERMANVI 578
Db 572 NQLSMDNKLRLSLKSLYTPAVDQYGNK---RSGLEPFLTNSLVDRNTDNTDNTAQNYS 628
QY 579 FTRAHSEVQTVIAKIIKAQINPKTDGLTFLDELKQAFKIYNEDMRQAKKYTQSNIPT 638
Db 629 FTRAHSEVQTVIAEIIKORIDPDSGLSPMDQLTEAFKIYNADQLKTDKEFTQNIIS 688
QY 639 AYALMLSNKDSITRLYYGDMYSDDGYMATKSPYYDAIDTLLKARIKYAAGQGDOKITYV 698
Db 689 TYATILTNKDTVPRVYGYDMYTDGQYMATKSLYDAIDTLLKSRIKYVSGGQTMKYM 748
QY 699 EGDKSHMDWYTGVLTSVRYGTGANEATOGSEATKTQGMVITSNPSKLQNQDKIV 758
Db 749 QGDSSMAADSYRGILTSVRYGTGANEATDAGTNETRQGIATVIESNPNLKLSSDQVVV 808
QY 759 NGAHAHKEYEPFLLLTKDGLTSYTSDA--AAKSLVRKTDKGELVFDASDIQGYLNPQV 817
Db 809 DMGIAHKNQAYRALLTTKGDITVYSDVSQSLRYTNSGQLIFNSSDIVGTANPQV 868
QY 818 SGYLAWVPVPGASDQVRAASNAKANATQGVYESSALDSQLIYEGFSNFQDFVTKSD 877
Db 869 SGYLAWVPVPGASDQVRAASNAKANATQGVYESSALDSQLIYEGFSNFQDFVTKSD 928
QY 878 YTNKIAQNVQIFKSGVTSPEMAPOYVSSSEGSFLDSIIQNGYAPEDRYDLAMSNNKY 937
Db 929 YANVOIANNITLYKSGWITNFEPFPQYRSSTDSFLDSIIQNGYAFTRDYDLGFNTPKY 988
QY 938 GSOQDMINAVKALHKSQIOVIADWPDQIYNLPKGEVVTATRVNDYGEYRKDSEIKNTLY 997
Db 989 GTVDQRTAKALHATGIKAMADWPDQIYNLPKGEVVAQVRVNSGIIYNQDSVINKTLY 1048
QY 998 AANTKSNKGDYQAKYGGAFSLAELAKYPSIFNRTQISNGKIDPSEKITAFAKAYFNGTN 1057
Db 1049 ASQIIGGG--EYQALYGGFELDEIKLYPALFEKNQISTGVPMDASEKIKENSAKYFNGTN 1107
QY 1058 ILGRGVYVVLKONADSKYFEL---KGNQTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQ 1114
Db 1108 IQCRGAYVYVKDWAYNEFKVSTSSNSVFLPKQLTNEESNTGFTSDGGMVYSTSGYQ 1167
QY 1115 AKNSFVQDAKGNWYFDNNHMYVGLQOLNGEVQVFLSNGVQLRESFLENAGSKNYFGH 1174
Db 1168 AKDFTIQDDKSNWYFDKNGYTYGFQTVNDDNYPFLPGLQDAILEDKSNVYFNG 1227
QY 1175 LGNRYNSGYSPDNDSKRYFDASGVMA--VGLKTI---NGNTQVFDQGVQVKGAWITG 1229
Db 1228 YGQAVDGYMYLANKT--WRYFDKNGVMAAGITTVTDQVQHIOVFDKNGIQVGTSVKD 1286
QY 1230 SDGKRYFDDGSGNMAVNRFPANDKNGDWYLYNSDGLYALGVQTINGKTYFYQDGKQIKG 1289
Db 1287 ADGKLYRFDSDGDMVNRFGENTDGTWYSYFGADGIATVGAQTISQKLFDFDADGQGIKG 1346
QY 1290 KIITD--NGKLYFLANSGLARNIATDQSNQWYFGSDGAVTGSQTTAGKLYFASDG 1348
Db 1347 KEATDKGKGVHYDANSSEMINTNRPEKLSDGSWAYFNKKNITVTGAQVINGQHLFFESNG 1406

QY 1349 KQVKG-SFTYNGKVVHYHADSGELQVNRPEADKQGNWYLYDSNGEALTGSQRINDQRVF 1407
Db 1407 NQVKGREYATDQKRYDYDADSGDMYTNRFERISDGSWAYFDANGVAVTGEQINGQQLY 1466
QY 1408 FTREGKQVKG 1417
Db 1467 FDANGHGVKG 1476

RESULT 11
AAU98038
ID AAU98038 standard; protein; 1475 AA.
XX
AC AAU98038;
XX
DT 27-AUG-2002 (first entry)
XX
DE S. mutans GTFB mutant I448V/D457N/D567T/D571K/K779Q/K1014T.
XX
KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture; mutant; mutein.
XX
OS Streptococcus mutans.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 448 /note= "wild-type Ile substituted by Val"
FT Misc-difference 457 /note= "wild-type Asp substituted by Asn"
FT Misc-difference 567 /note= "wild-type Asp substituted by Thr"
FT Misc-difference 571 /note= "wild-type Asp substituted by Lys"
FT Misc-difference 779 /note= "wild-type Lys substituted by Gln"
FT Misc-difference 1014 /note= "wild-type Lys substituted by Thr"
XX
PN US2002031826-A1.
XX
PD 14-MAR-2002.
XX
PF 19-DEC-2000; 2000US-00740274.
XX
PR 07-JUN-1995; 95US-00478704.
PR 07-JUN-1995; 95US-00482711.
PR 07-JUN-1995; 95US-00485243.
PR 16-JAN-1998; 98US-00007999.
PR 16-JAN-1998; 98US-00008172.
PR 20-JAN-1998; 98US-00009620.
PR 11-DEC-1998; 98US-00210361.
XX
PA (NICH/) NICHOLS S E.
XX
PI Nichols SE;
XX
DR WPI; 2002-414332/44.
XX
PT Glucosyltransferase B or D protein useful for producing a glucan useful
PT as substitutes for and additions to modified starch and latexes in paper
PT manufacture, comprises mutations in specific positions.
XX
PS Claim 36; Page; 44pp; English.
XX
CC The invention an isolated protein comprising a glucosyltransferase (GTF)
CC B polypeptide having changes at position from I448V, D457N, D567T,
CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,
CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y109A/Y170A/Y171A, and K779Q or a
CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
```


complementary polynucleotide, a ribonucleic acid sequence encoding the GTF mutant, an expression cassette comprising the polynucleotide operably linked to a promoter, a vector comprising the expression cassette, host cell introduced with the vector, a transgenic plant comprising the vector, a seed or tuber from the transgenic plant, a paper sizing and/or coating composition comprising a glucan produced in a plant transformed with a gene encoding the mutant GTF, wild type or, starch, a latex, thermoplastic molecule or their combinations or glucan and starch where the glucan is produced in the amyloplast and/or vacuole or a maize line deficient in starch biosynthesis, transformed with a gene encoding a glucosyltransferase B or D enzyme, wild-type or mutant and a paper comprising the glucan (paper sizing/coating agent). The vector is useful for producing a glucan in a plant. The method comprises transforming a plant cell with the vector, growing the plant cell under plant growing conditions to produce a regenerated plant and inducing expression of the polynucleotide for a time sufficient to produce the glucan in the regenerated plant, where the vector contains a transit sequence from ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and chlorophyll AB binding protein to produce a transgenic plant, and glucan is produced in the amyloplast of potato or the vacuole of sugar beet. Glucans are useful as substitutes for or additions to modified starch and latexes in paper manufacture. Unlike prior art techniques, which require input materials that produce chemical effluents, paper manufacture utilising the glucan produced by GTF, which utilises biologically produced input materials, is more cost-effective and environmentally friendly. Moreover, Glucans also exhibit thermoplastic properties and impart gloss to the paper during coating step. The present sequence represents a GTFB mutant of the invention. Note: The present sequence is not shown in the specification but was created by the indexer using the GTFB sequence appearing as AAU98027 and the information in claim 36

Sequence 1475 AA;

Query Match	51.5%;	Score	3845.5;	DB 5;	Length	1475;			
Best Local Similarity	53.3%;	Pred.	No. 8.9e-191;						
Matches	768;	Conservative	227;	Mismatches	392;	Indels	53;	Gaps	22;
1	METKERYKMHKVKKGWTVAVASGLITLGTITLTTSSVSAETEQQTSDKVVTQKSEDDRAA	60	:	:	:	:	:	:	:
1	MDKKVRYKLKKVKKGWTVSVASAVWTL-TTLUSGGLVKADSNEKS-----QISNDSNTS	54	:	:	:	:	:	:	:
61	SESSOTDAPKTKQAQTEQTQAQSQAN-VADTSTSTIKETPSSQNIITQAANSDDDKTVTNTKS	119	:	:	:	:	:	:	:
55	VVTANEESNVITEATSKQEAASQTNHVTVTSSSTSVVNPKEVSVNPVTUGETASNGEK	114	:	:	:	:	:	:	:
120	EEAQTSEERTKQSEEAQTTPASSQALTQAKAEITKQRTAAQENKQNPVDLAAIPNVKQIDG	179	:	:	:	:	:	:	:
115	LQNTQTTV-DKTSEAAANNISKQT-TEADTDVIDDSNAA-----NLQILEKLPNVKEIDG	167	:	:	:	:	:	:	:
180	KYYIGSGQPKKNFALTNNKVLVPDKNTGALTQTS-QYQFKQGLTKLNNDYTPHNQIV	238	:	:	:	:	:	:	:
168	KYYVDNNGKVRNFTLIADGKILHPDE-TGAYTQTSIDTVNKOIVTTRSNLYKKYNQVY	226	:	:	:	:	:	:	:
239	NPENTSLETIDNVYTADSNMYRPKDILQNGKTTTASSESLRPLLMWSPDKQQTAYLNY	298	:	:	:	:	:	:	:
227	DRSAQSFHVHYLTAESHYRKYILKDGKTTWQSTEKDFRPLMTWPDQSTQRQVNY	286	:	:	:	:	:	:	:
299	MNQQLGTGENTYADSSQSLSLAAQTQVVKIETKISQTOQTOWLRDIINSFVKTPQNNW	358	:	:	:	:	:	:	:
287	MNAQ-IQINKTYDDTSNQLNLNAAATIQAQTEAKITLTKNTDWLRQTISAFVKTSQAWN	345	:	:	:	:	:	:	:
359	SQTESDTSAGEKHILQGGALLYSNDSK-TAYANSOYRLNRRPTSQTK--PKYFEDNSS	415	:	:	:	:	:	:	:
346	SDESEKPFD-----DHILQNGAVLVDFNEGKLPYANSNYRILNRRPTQTQKKDPRYTADNTI	401	:	:	:	:	:	:	:
416	GGYDFLLANDIDNSNPVQAQQLNHLWLMNYGSTVANDPEANFQGVRVDAVDNVDNADLL	475	:	:	:	:	:	:	:
402	GGYEFLLANDVNSNPVQAQQLNHLWLFNFGNIYANDPDANFDSVRVDAVDNVDNADLL	461	:	:	:	:	:	:	:
476	QIASDYLKAHYGVQVDSKEKNAINHLSILEAWSNDNDPOYNKQTKGAQLPDNKLRLSLLYAL	535	:	:	:	:	:	:	:
462	QIAGDYLKAAKGIHKNDKAANDHLSLILEAWSNDNTPYLHDDGDNIMNDKRLSLLSFSL	521	:	:	:	:	:	:	:

Qy	536	TRPLEKDA	SNKNKEIRSGLEPVIITNSLNRSABGKNRGMANYIPRAHDSDEVQTIAKII	5395
Dd	522	AKPLNQ-----	RSGMNPILTNSLVNRDDNAETAAPVSPSFIRAHDSEVOQLTAKII	573
Qy	596	KAOINPKTDGLTFETLDELQA	FKIYNEDMQAKKYTQSNIPTAYALMLSNKDSTRLYY	655
Dd	574	KAEINPNVGYS	TMEIEIKAFSIYKNDLUATSEKKYTHNTALSYALLLTNKS	633
Qy	656	GDMYSDDQYMA	TSKSPYDAIDTLTKARIKYAAGGDMMKITVYEGDKSHMDWDYTGVLTLS	715
Dd	634	GDMFTDGGYMA	HKTINYEALETLLKARIKYVSGQAMRNQQVGNSE-----IITS	684
Qy	716	VRYGTGANEAT	DGGSSEATKTQGMAVITSSNPSSIKLNDXKIVIMNGAAHKXQYRP	775
Dd	685	VRYGKGALKAT	DGDRITRTTSYGAVIIEGNNPSURLKASDRVVVMGAAHKXQYRP	744
Qy	776	TKDGLTSYSDAA	KSILYRKTNDEKGLVFDPASDIQGYLPNPQVSGYLAVVPVCASDNQDV	835
Dd	745	TDNGIKRAYHS	DQBAAGLVRYTYNDRGELIFTAADIQGANPOVSIGYLGWVPVGA	: : : : : 801
Qy	836	RVAASNKANAT	QGV--YESSALDSOLIYEFGSFNFQDFVTUKDSDYTNKKIAQNVLFKSW	893
Dd	802	KMFALRLARPH	QOMASVHQNAALDSRWMPESGFNFQAFATKEBEYTNVVIKANNVDKFAEW	861
Qy	894	GVTSSFENAPQ	YVSESDEGSFLDSIIIONGAFEDREDYDLAWSKNKYKGSOODMINAVKALHKS	953
Dd	862	GVTDFENAPQ	YVSVSTDGSLFDSVIQNGYAFTDRYDLGISKNPKYGTADDDLKAIALHKS	921
Qy	954	GIOVIADWP	QDLYNLPGKEVVTATRVDNYGEVRKDSSEIKNTLIYAANTKNSGXDYOAKYG	1013
Dd	922	GIKWMADWP	QMYAPEKEVVTATRVDKYGTPVAGSQIKNTLIYVVDGKSSGKDOQAQYG	981
Qy	1014	GAFELSELA	KYPISEIFNRTQISNGKKIDPSEKITAWKAKYPNGTNILGRGVGYVLKDNASD	1073
Dd	982	GAFLEELOAKY	PFLFARKQISTGVPMDDPSVKIITOWSAKYENGNTILCRGAGYVLKDAQTN	1041
Qy	1074	KYPELKGNQ--	TYLPKOMTNKEASTGFVNDGNGMTFYSTSGYOAKNSFVODAKGNWYFD	1131
Dd	1042	TYFNISDNKE	INFPLPKTLLNQVGFSGYDGKGYVYSTSGYOAKNTFISEG-DKWYFFD	1100
Qy	1132	NNGHMVYGLQ	LANGEVOYFLSNSGVQLRESFLENADGSKNYFHGLGNRYSNGYYSFONDNSK	1191
Dd	1101	NNGYVWTGA	SGINGVNYYFLSNGLQLRDALTKNEDGTIAYYNGDRREYENGYQF-MSGV	1159
Qy	1192	WRYPDASGV	MVAVGLKTNGNTQYFDQGYOVKGAWITGSGDKKRYFPDGGSNMAVNRPAN	1251
Dd	1160	WRHFN-NGEM	SVGLTWIDGQVQVFDGMYOAKGKFVTTADGKIRYFDKQSGNMYYRRNFIE	1218
Qy	1252	DKNGDWYLL	NSDGIALGVQTINGKTYFYQDCGKOIKGLIITD-NGKLKYFLANSGLAR	1310
Dd	1219	NEEGKWLIV	GEDGAAVTGSGTINQHLLYFRANGVQVKGFEVTDHHGRISITYDNGSGHQIR	1278
Qy	1311	NIFATDSQNN	WYFGSDGAVTGSQTIAGKLYFASDGKVKGVSFVTYN-GKHVHYHAD	1369
Dd	1279	NRFVRNAQ	QGWQFYFDNNNGYAVTGARTINGQLLYFRANGVQVKGFEVTDYGRISYYD	1338
Qy	1370	GELOVNRPE	ADKGNWYLLDSNGEALTGSGQRINDORVFTFREKGQVKGVDVAYDERLLLVY	1429
Dd	1339	GDOIQRNRF	VRNAQGWQFYFDNNGYAVTGARTINGQHLLYFRANGVQVKGFEVTDHRGRISY	1398
 RESULT 12				
AAU98037	ID	AAU98037 standard; protein; 1475 AA.		
XX	AC	AAU98037;		
XX	CC	(first entry)		
DT	27-AUG-2002			
XX	S. mutans	glucosyltransferase GTFB mutant D567T/D571K/K1014T.		
DE	Glucosyltransferase; GTFB;	transgenic plant; paper sizing;		
XX	XX			

QY 239 NFENTSLETIDNYTADSWTRPKDILKNGKWTWTSSESDLRPLLMWWPKQTOIAYLNY 298
Db 227 DRSAQSPGTHVHYLTAEBSWTRPKYILKDGKWTOSTEKDFRPLMTWPDQETORQYVY 286
QY 299 MNOQGLGTGNYTADSQESLNLAQOTVQVKIETKISOTOQWLRDIINSFVKTOPNWN 358
Db 287 MNAQ-LGINKTYDDTSNQLQNTAATIQAKIEAKITILKNTDWLRQTSIAFVKTSQAWN 345
QY 359 SOTESDTSAGEKHOGLQGALLYNSDK-TAYANSYRLNRTPTSTQK--PKYFEDNSS 415
Db 346 SDSEKPPD---DHLQNGAVLYDNEGKLTTPYANSYRLNRTPTNQTKKDPRTADNTI 401
QY 416 GGYDFLLANDIDNSNPVQAEQLNLWHLNYGSIIVANDPEANPDGVVRDADVNADLL 475
Db 402 GGYEFLANDVDNSNPVQAEQLNLWHLNFGNIYANDPDANFDSIRVDADVNADLL 461
QY 476 QIASDYLAHYGVDKSEKNAIHLISLEAUSDNDPQNKDTKGAQLPIDNKLRLSLAL 535
Db 462 QIAGDYLYKAAKGTHKNDKAANDHLSLEAUSDNDTPYLHDGDNMINDNKLRLSLFL 521
QY 536 TRPLEKXDAKNKIBIRSGLEPVIITNSLNRSAEGKNSERMANYIFIRAHDSVQTVIAKII 595
Db 522 AKPLNQ-----RSCMNPILINSLVNRDDNAETAAPVPSYFIRAHDSVQTVIAKII 573
QY 596 KAOINPKTDGLTFLDELKQAFKIYEDMRQAKKYYTQSNIPYAYALMLSNKDSITRLYY 655
Db 574 KAEINPNVGVSTBEIKKAFIYKNDLLATEKKYTHYNTALSYALLTNKSVPRVY 633
QY 656 GDMYSDGQVMATKSPYDAIDTLKARIKAYAGGQDMKITYEGDKSHMDWDTYGVILTS 715
Db 634 GDMFTDGGYMAHKTINYEATIELLKARIKYVSGQAMRQQVGNSE-----IITS 694
QY 716 VRYGTGANEATDQGSBATKTQGMVITSNPISLKLQNQNDKVIYVMGAHKNQYRPLLLT 775
Db 685 VRYGKALKATDGTDRTRTSGVAVIEGNNPSRLKASDRVNVNMGAAHKNQYRPLLLT 744
QY 776 TKDGLTSYSDAAKSLYKNTNDKGLVFDASDIQGYLPQVSGYLAVVPGVQASQNDV 835
Db 745 TONGIKAYHSDQEAAGLVRYTNDRGELIETAAIDIKGYANPQVSGYLGVVVPVGA--LI 801
QY 836 RVAASNAKANATQV--YESSSALDSQLIYEGESNFQDFVTKDSYNTKKIAQNVQPKSW 893
Db 802 KMFALKLAPHOQMAVSHQNAALDSRMFEGFSNFOAFATKBEETNVVIAKNVDKFAEW 861
QY 894 GVTSEMAPQYVSSEDSGLDSIIQNGYAFEDRYDLAMSKNNKYQSQDMINAVKALHS 953
Db 862 GVTDFEMAPQYVSSTDSGLDSVIQNGYAFEDRYDLGISKPNKYGTADDLVKAIALHSK 921
QY 954 GIOVDADVPDQIYNLPGEVVTATRVNDYGEYRKDSEIKNTLYAANTKSNKGQYQAKYG 1013
Db 922 GIKWADWVPDQMYAPPEKEVVTATRVKYGTVPVAGSQIKNTLYVVDGKSSGKQQAKYG 981
QY 1014 GAFLESLAKYPSIFNRTQISNGKIDPEKITAWAKYFNGNTNILGRGVGVYLNKNASD 1073
Db 982 GAFLEELQAKYFELFARKQISTGVPMDSVKIKQWSAKYFNGNTNILGRGAGYVYLKQDQATN 1041
QY 1074 KYFELKNGO--TYLPMQNTNKBASTGFVNDGNGMTFYSTGYQAKNSFVODAKGNWYFD 1131
Db 1042 TYFNTSDNKEINFLPKTLLNQDSQVGSFVDGKGYVYSTGQAKNTFISEG-DKWYIFD 1100
QY 1132 NNGHMYGLQQLNGEVQYFSLNGVQLRESFLENADGSKNYFGLGNRYNGYYSFNDNSK 1191
Db 1101 NNGYMTGAQSIINGVNYVYFSLNGLQRLDAILKNEDGTAYAYGNDGRYENGYQF-MSGV 1159
QY 1192 WRYFDASGVMAVGLKTINGTQYFDQGGYQVKGAMITGSDGKKRYPDGSGMNAVNRPAN 1251
Db 1160 WRHFN-NGEMSVGLTVIDQGVYFDEMGYQAKGKFTTADGKIRYPDQSGNMYNRFT 1218
QY 1252 DKNQDWWYNSDIALGVQTINGKTYTFYQDGKQIKGKIITD-NGKLKYFLANSGLAR 1310
Db 1219 NEEGKWLIGEDGAATVGSQTINGQHLYFRANGVQVKGFEVTDHHRGISYDNGSDQIR 1278
QY 1311 NIFATDSQNNWYFSGDGVAVTGSQTIAGKLYFASDQKVGKSFVTYN-GKVHYHADS 1369

Db 1279 NRVFNAQSQWFFYDNNGYAVTGARTINGQLLYFRANGVQVKGFEVTDYGRISYDNGS 1338
QY 1370 GELQVNRFEADKGNWVYLDNNGEALTGSQRINDORVFFTRGKQVKGVDVAYDERRLLVY 1429
Db 1339 GDQIRNRFVNAQSQWFFYDNNGYAVTGARTINGQHLYFRANGVQVKGFEVTDYGRISY 1398

RESULT 14

AAU98035
ID AAU98035 standard; protein; 1475 AA.
XX AAU98035;
XX 27-AUG-2002 (first entry)
XX S. mutans glucosyltransferase GTFB mutant D457N/D571K.
DE XX Glucosyltransferase; GTFB; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture; mutant; mutain.
XX Streptococcus mutans.
OS Synthetic.
XX Key Location/Qualifiers
FH Misc-difference 457 /note= "Wild-type Asp substituted by Asn"
FT Misc-difference 571 /note= "Wild-type Asp substituted by Lys"
FT
XX US2002031826-A1.
XX 14-MAR-2002.
XX 19-DEC-2000; 2000US-00740274.
XX 07-JUN-1995; 95US-00478704.
PR 07-JUN-1995; 95US-00482711.
PR 07-JUN-1995; 95US-00485243.
PR 16-JAN-1998; 98US-00007999.
PR 16-JAN-1998; 98US-00008172.
PR 20-JAN-1998; 98US-00009620.
PR 11-DEC-1998; 98US-00210361.
XX (NICH/) NICHOLS S E.
XX Nichols SE;
XX WPI; 2002-414332/44.
XX Glucosyltransferase B or D protein useful for producing a glucan useful
as substitutes for and additions to modified starch and latexes in paper
manufacture, comprises mutations in specific positions.
PS Claim 36; Page; 44pp; English.
XX The invention an isolated protein comprising a glucosyltransferase (GTF)
B polypeptide having changes at position from 1448V, D457N, D567T,
CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/K1014T,
CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
CC complementary polynucleotide, a ribonucleic acid sequence encoding the
CC GTF mutant, an expression cassette comprising the polynucleotide operably
CC linked to a promoter, a vector comprising the expression cassette, host
CC cell introduced with the vector, a transgenic plant comprising the
CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
CC coating composition comprising a glucan produced in a plant transformed
CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
CC thermoplastic molecule or their combinations or glucan and starch where
CC the glucan is produced in the amyloplast and/or vacuole or a maize line

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 19:39:30 ; Search time 31.3088 Seconds
(without alignments)
3776.130 Million cell updates/sec

Title: US-10-797-821-36
Perfect score: 7462
Sequence: 1 METKRRYKMHVKYKHVTVTA.....EGKQVKGDVAYDERLLVYR 1430

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5_COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/6_COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/H_COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/pCTUS_COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/RE_COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7462	100.0	1430	2	US-09-008-172-2
2	7462	100.0	1430	2	US-09-210-361-6
3	7462	100.0	1430	2	US-09-740-274-6
4	3817.5	51.2	1475	2	US-09-007-999-2
5	3817.5	51.2	1475	2	US-09-210-361-2
6	3817.5	51.2	1475	2	US-09-740-274-2
7	3788.5	50.8	1577	1	US-08-793-824-2
8	3591.5	48.1	1375	2	US-09-210-361-4
9	3591.5	48.1	1375	2	US-09-740-274-4
10	2808	37.6	545	2	US-09-604-957-4
11	2780	37.3	545	2	US-09-995-749A-10
12	2699.5	36.2	2057	2	US-09-499-203-2
13	2446	32.8	1278	2	US-09-604-957-3
14	2446	32.8	1781	2	US-09-995-749A-2
15	1534	20.6	523	2	US-09-604-957-5
16	1532.5	20.5	522	2	US-09-995-749A-11
17	1400	18.8	535	2	US-09-604-957-7
18	1400	18.8	535	2	US-09-995-749A-13
19	1308.5	17.5	584	2	US-09-604-957-6
20	1306.5	17.5	584	2	US-09-995-749A-12
21	595.5	8.0	349	2	US-09-008-620-2
22	453	6.1	2710	1	US-08-480-604A-6
23	453	6.1	2710	1	US-08-405-496A-6
24	453	6.1	2710	2	US-08-915-136-6
25	453	6.1	2710	2	US-08-957-310-6
26	453	6.1	2710	2	US-10-011-366-6
27	453	6.1	2710	2	US-09-084-517-6

28	390.5	5.2	1231	2	US-08-714-741-41	Sequence 41, Appl
29	387	5.2	2366	1	US-08-480-604A-10	Sequence 10, Appl
30	387	5.2	2366	1	US-08-405-496A-10	Sequence 10, Appl
31	387	5.2	2366	2	US-08-915-136-10	Sequence 10, Appl
32	387	5.2	2366	2	US-08-957-310-10	Sequence 10, Appl
33	387	5.2	2366	2	US-10-011-366-10	Sequence 10, Appl
34	387	5.2	2366	2	US-09-084-517-10	Sequence 10, Appl
35	363	4.9	811	1	US-08-480-604A-7	Sequence 7, Appl
36	363	4.9	811	1	US-08-405-496A-7	Sequence 7, Appl
37	363	4.9	811	2	US-08-915-136-7	Sequence 7, Appl
38	363	4.9	811	2	US-08-957-310-7	Sequence 7, Appl
39	363	4.9	811	2	US-10-011-366-7	Sequence 7, Appl
40	363	4.9	811	2	US-09-084-517-7	Sequence 7, Appl
41	363	4.9	812	1	US-08-480-604A-29	Sequence 29, Appl
42	363	4.9	812	1	US-08-915-136-29	Sequence 29, Appl
43	363	4.9	812	2	US-09-084-517-29	Sequence 29, Appl
44	359	4.8	866	2	US-09-545-773-2	Sequence 2, Appl
45	359	4.8	866	2	US-10-222-038-2	Sequence 2, Appl

RESULT 1
US-09-008-172-2
; Sequence 2, Application US/09008172
; Patent No. 6127602
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0358D
; CURRENT APPLICATION NUMBER: US/09/008,172
; CURRENT FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-008-172-2

Query Match	100.0%	Score 7462;	DB 2;	Length 1430;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 1430;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	METKRRYKMHVKYKHVTVAVASGLITLGTTLGSSVSAETEQOTS	Sequence 6, Appl
Db	1	METKRRYKMHVKYKHVTVAVASGLITLGTTLGSSVSAETEQOTS	Sequence 6, Appl
QY	61	SESSOTDAPKTKQAQTEQTSQANVADTSTSTITKTPSQNTTQANSDDKT	Sequence 2, Appl
Db	61	SESSOTDAPKTKQAQTEQTSQANVADTSTSTITKTPSQNTTQANSDDKT	Sequence 2, Appl
QY	121	EAQTSSEETKQSEEAQTASSOALTQAKAELTKQRTAOKENKPNVDLAAI	Sequence 10, Appl
Db	121	EAQTSSEETKQSEEAQTASSOALTQAKAELTKQRTAOKENKPNVDLAAI	Sequence 10, Appl
QY	181	YYYIGSDQPKKNFALTVNNKLVYFDKNTGALTDTSTQYQKGLTKLNNDY	Sequence 240
Db	181	YYYIGSDQPKKNFALTVNNKLVYFDKNTGALTDTSTQYQKGLTKLNNDY	Sequence 240
QY	241	ENTSLTIDTVYADSWYRPKDILKNGKWTWASSEDRLPLMSWPKQTOIAY	Sequence 300
Db	241	ENTSLTIDTVYADSWYRPKDILKNGKWTWASSEDRLPLMSWPKQTOIAY	Sequence 300
QY	301	QOGLGTGENYADSSQESLNLAQTVQKIETKISQTOQTQWLRLDIINSFV	Sequence 360
Db	301	QOGLGTGENYADSSQESLNLAQTVQKIETKISQTOQTQWLRLDIINSFV	Sequence 360
QY	361	TESDTSAGEKDLQGGALLYNSDKTAVANSYRLNLTPTTSQTKPKYFED	Sequence 420

ALIGNMENTS

Db 361 TESDTISAGEKDHLLQGALLYSNSDKTAYANSYRLNRTPTSTQTKPKYFEDNSSGGYDF 420
QY 421 LLANDIDNSNPVQAEQLNWLHLYLMNYGSIIVANDPEANFDGVRVDVAVNNADLLQIASD 480
Db 421 LLANDIDNSNPVQAEQLNWLHLYLMNYGSIIVANDPEANFDGVRVDVAVNNADLLQIASD 480
QY 481 YLKAHYGVDSKSEKNAINHLSILEANSNDNDPQYNKDTKGALPDKLRLSLLYALTRPLE 540
Db 481 YLKAHYGVDSKSEKNAINHLSILEANSNDNDPQYNKDTKGALPDKLRLSLLYALTRPLE 540
QY 541 KDAASKNEIRSGLEPVTITNSLNRSAEGKNSEMANIYIFIRAHDSVQTVIAKIIKAQIN 600
Db 541 KDAASKNEIRSGLEPVTITNSLNRSAEGKNSEMANIYIFIRAHDSVQTVIAKIIKAQIN 600
QY 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYYTQSNIPTAYALMLSNKDSITRLYYGDMYS 660
Db 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYYTQSNIPTAYALMLSNKDSITRLYYGDMYS 660
QY 661 DDQYMATKSPYDAIDTLLKARIKYAAGQDMKITYVEGDKSHMDWDYTGVLTSVRYGT 720
Db 661 DDQYMATKSPYDAIDTLLKARIKYAAGQDMKITYVEGDKSHMDWDYTGVLTSVRYGT 720
QY 721 GANEATDQGSSEATKTQGMVITSNPSLKNQNDKVI VNMGAHKNQOEYRPLLLTTKQGL 780
Db 721 GANEATDQGSSEATKTQGMVITSNPSLKNQNDKVI VNMGAHKNQOEYRPLLLTTKQGL 780
QY 781 TSYTSDDAAAKSLYRKTNDKGELVFDASDIQGYLNPQVSGYLAVVVPVGASDNQDVRVAAS 840
Db 781 TSYTSDDAAAKSLYRKTNDKGELVFDASDIQGYLNPQVSGYLAVVVPVGASDNQDVRVAAS 840
QY 841 NKANATQGYYESSALDSQIIVEGFSNFDQFVTKDSDYTNKKIAQNVQLFKSGWVTSFPM 900
Db 841 NKANATQGYYESSALDSQIIVEGFSNFDQFVTKDSDYTNKKIAQNVQLFKSGWVTSFPM 900
QY 901 APOYVSSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSCQDMINAVKALHKSIGIOVIAD 960
Db 901 APOYVSSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSCQDMINAVKALHKSIGIOVIAD 960
QY 961 WVPDQIYNLPKGEVVTATRVNDYGEYRKDSEIKNTLYAANTKSNKDYQAKYGGAFSLSEL 1020
Db 961 WVPDQIYNLPKGEVVTATRVNDYGEYRKDSEIKNTLYAANTKSNKDYQAKYGGAFSLSEL 1020
QY 1021 AAKYPSIFNRTQISNGKKIDPSEKITAWKAKYFNGTNIILGRGVGYVLKDNASDKYFELKG 1080
Db 1021 AAKYPSIFNRTQISNGKKIDPSEKITAWKAKYFNGTNIILGRGVGYVLKDNASDKYFELKG 1080
QY 1081 NOTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYIFDNNGHMVYGL 1140
Db 1081 NOTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYIFDNNGHMVYGL 1140
QY 1141 QQLNGEVQYFLSNGVQLRESFLENADGSKNYFGLHGNRYSNGYYSFNDNSKWRYPDASGV 1200
Db 1141 QQLNGEVQYFLSNGVQLRESFLENADGSKNYFGLHGNRYSNGYYSFNDNSKWRYPDASGV 1200
QY 1201 MAVGLKTINGNTQYFPDQGYQVKGAWITGSDGKKRYFDDGSGNMAVNRFANDKNGDWYYL 1260
Db 1201 MAVGLKTINGNTQYFPDQGYQVKGAWITGSDGKKRYFDDGSGNMAVNRFANDKNGDWYYL 1260
QY 1261 NSDGIALGVQTINGNTYTFGQDGKQIKGKITDNGKLYFLANSSELARNIPATDSQNN 1320
Db 1261 NSDGIALGVQTINGNTYTFGQDGKQIKGKITDNGKLYFLANSSELARNIPATDSQNN 1320
QY 1321 WYVFGSDGVAVTGSQTIAGKCLYFASDGQVKGVSFVYTKNGKVHYHADSGELQVNRFEAD 1380
Db 1321 WYVFGSDGVAVTGSQTIAGKCLYFASDGQVKGVSFVYTKNGKVHYHADSGELQVNRFEAD 1380
QY 1381 KQGNWYILDSNGEALTGSRINDQRVFFTRREGQVKGVDVAYDBERRLLLVYR 1430
Db 1381 KQGNWYILDSNGEALTGSRINDQRVFFTRREGQVKGVDVAYDBERRLLLVYR 1430

US-09-210-361-6
; Sequence 6, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; EARLIER FILING DATE: 1998-12-11
; EARLIER FILING DATE: 1998-01-16
; EARLIER FILING DATE: 1995-06-07
; EARLIER FILING DATE: 1995-06-07
; EARLIER FILING DATE: 1998-01-16
; EARLIER FILING DATE: 1998-01-20
; EARLIER FILING DATE: 1995-06-07
; EARLIER FILING DATE: 1995-06-07
; EARLIER FILING DATE: 1998-01-16
; EARLIER FILING DATE: 1998-01-20
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-210-361-6

Query Match 100.0%; Score 7462; DB 2; Length 1430;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 METKERYKMHKVKHWTAVASGLITLGTTLTGGSSVSAETEQSDSKVVTQKSEDDKAA 60
Db 1 METKERYKMHKVKHWTAVASGLITLGTTLTGGSSVSAETEQSDSKVVTQKSEDDKAA 60
QY 61 SSSOTDAPKTKQAOTEQTAOSQANVADTSTSIKETPSQNIITQANSDDKTVTNTKSE 120
Db 61 SSSOTDAPKTKQAOTEQTAOSQANVADTSTSIKETPSQNIITQANSDDKTVTNTKSE 120
QY 121 EAQTSEERTKQSEEAQTASSQALTKQAELTKQRTAQENKNPVDLAAIPNVKQIDGK 180
Db 121 EAQTSEERTKQSEEAQTASSQALTKQAELTKQRTAQENKNPVDLAAIPNVKQIDGK 180
QY 181 YYYIGSDGQPKKNFALTNNKVLTFDKNTGALTDTTSQYQFKGLTKLNNDYTPHNQIVNF 240
Db 181 YYYIGSDGQPKKNFALTNNKVLTFDKNTGALTDTTSQYQFKGLTKLNNDYTPHNQIVNF 240
QY 241 ENTSLETIDNYVTADSWTRPKDILKNGKTTWTASSEDRLPILLMSWVPDKQTQIAYLNYN 300
Db 241 ENTSLETIDNYVTADSWTRPKDILKNGKTTWTASSEDRLPILLMSWVPDKQTQIAYLNYN 300
QY 301 QQGLGTGENYADSSQESLNLAQTVQVKIETKISQTOQOTWLRDILINSFVKTPQPNNSQ 360
Db 301 QQGLGTGENYADSSQESLNLAQTVQVKIETKISQTOQOTWLRDILINSFVKTPQPNNSQ 360
QY 361 TESDTSAGEKQHLQGGALLYSNSDKTAYANSYRLNRTPTSTQTKPKYFEDNSSGGYDF 420
Db 361 TESDTSAGEKQHLQGGALLYSNSDKTAYANSYRLNRTPTSTQTKPKYFEDNSSGGYDF 420
QY 421 LLANDIDNSNPVQAEQLNWLHLYLMNYGSIIVANDPEANFDGVRVDVAVNNADLLQIASD 480
Db 421 LLANDIDNSNPVQAEQLNWLHLYLMNYGSIIVANDPEANFDGVRVDVAVNNADLLQIASD 480
QY 481 YLKAHYGVDSKSEKNAINHLSILEANSNDNDPQYNKDTKGALPDKLRLSLLYALTRPLE 540
Db 481 YLKAHYGVDSKSEKNAINHLSILEANSNDNDPQYNKDTKGALPDKLRLSLLYALTRPLE 540
QY 541 KDAASKNEIRSGLEPVTITNSLNRSAEGKNSEMANIYIFIRAHDSVQTVIAKIIKAQIN 600
Db 541 KDAASKNEIRSGLEPVTITNSLNRSAEGKNSEMANIYIFIRAHDSVQTVIAKIIKAQIN 600

601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYYTOSNIPTAVALMSNKDSITRLYYGDMYS 660
601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYYTOSNIPTAVALMSNKDSITRLYYGDMYS 660
661 DDGOYMATKSPYDAIDTLKAKIYAAGQDMKITYVEGDKSHMDWDYTGVLTSVRYGT 720
661 DDGOYMATKSPYDAIDTLKAKIYAAGQDMKITYVEGDKSHMDWDYTGVLTSVRYGT 720
721 GANEATDQGSSEATKTQGMVITSNPSLKLNDKVIYNMGAHKNQOEYRPLLLTKDGL 780
721 GANEATDQGSSEATKTQGMVITSNPSLKLNDKVIYNMGAHKNQOEYRPLLLTKDGL 780
781 TSYTSDAAKSLYRKTNDSKELVFDASDIQYLNPOVSGYLAVVPGASDNQDVRVAAS 840
781 TSYTSDAAKSLYRKTNDSKELVFDASDIQYLNPOVSGYLAVVPGASDNQDVRVAAS 840
841 NKANATGOVYESSALDSQIYEGFSNFQDFVTNKSDYTNKKTIAQNVLPKSGVTSFEM 900
841 NKANATGOVYESSALDSQIYEGFSNFQDFVTNKSDYTNKKTIAQNVLPKSGVTSFEM 900
901 APOVYSSDGSFLDSIIQNGYAFEDRYDLAMSKNKKYGSQODMINAVKALHKSIGIOVIAD 960
901 APOVYSSDGSFLDSIIQNGYAFEDRYDLAMSKNKKYGSQODMINAVKALHKSIGIOVIAD 960
961 WVPDQIYNLPKGEVVTATRVNDYGEYRKDSEIKNTLYAANTKSNKGDKYQAKYGGAFSEL 1020
961 WVPDQIYNLPKGEVVTATRVNDYGEYRKDSEIKNTLYAANTKSNKGDKYQAKYGGAFSEL 1020
1021 AAKYPSIFNRTQISNGKKIDPSEKITAWEKAYFNGTNILGRGVYVLDKNSDKYFELKG 1080
1021 AAKYPSIFNRTQISNGKKIDPSEKITAWEKAYFNGTNILGRGVYVLDKNSDKYFELKG 1080
1081 NOTYLPKQMTKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYFDNNGHMYGL 1140
1081 NOTYLPKQMTKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYFDNNGHMYGL 1140
1141 QOLNGEVQYFSLNGVQLRESFLENADGSKNYFGLHGNRYSGYSPDNDSKWRYPDASGV 1200
1141 QOLNGEVQYFSLNGVQLRESFLENADGSKNYFGLHGNRYSGYSPDNDSKWRYPDASGV 1200
1201 MAVGLKTNGNTQYFDQDGYQVKGAWITGSDGKKRYFDDGSGNMVNRFANDKNGDWYYL 1260
1201 MAVGLKTNGNTQYFDQDGYQVKGAWITGSDGKKRYFDDGSGNMVNRFANDKNGDWYYL 1260
1261 NSDGIALVGQTINGKTYTFQDQKQIKGKIITDNGKLYFLANSSELARNIFATDSQNN 1320
1261 NSDGIALVGQTINGKTYTFQDQKQIKGKIITDNGKLYFLANSSELARNIFATDSQNN 1320
1321 WYFSGDGVAVTGSQTIAGKLYFASDGKQVKGSFVYNGKVHYHADSGELQVNRFEAD 1380
1321 WYFSGDGVAVTGSQTIAGKLYFASDGKQVKGSFVYNGKVHYHADSGELQVNRFEAD 1380
1381 KQGNWYILDSNGEALTGSRINDQRVFFTRGQKQVGDVAYDERLLVYR 1430
1381 KQGNWYILDSNGEALTGSRINDQRVFFTRGQKQVGDVAYDERLLVYR 1430

RESULT 3

US-09-740-274-6
; Sequence 6, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704

; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-6

Query Match 100.0%; Score 7462; DB 2; Length 1430;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 METKRYKHKVKKHVVTVAVASGLITLGTTLGSSVSAETEQTSDKVVTKSEDDKAA 60
DB 1 METKRYKHKVKKHVVTVAVASGLITLGTTLGSSVSAETEQTSDKVVTKSEDDKAA 60
QY 61 SESSQTDAPKTQAOETEQTQAOQANVADTSTITKETPSQNTTQANSDDKVTWTKSE 120
DB 61 SESSQTDAPKTQAOETEQTQAOQANVADTSTITKETPSQNTTQANSDDKVTWTKSE 120
QY 121 EAQTSERTKQSEEAQTASSQALTOAKAELTKQRTAAQAKNPNVDLAAIPNVKQIDGK 180
DB 121 EAQTSERTKQSEEAQTASSQALTOAKAELTKQRTAAQAKNPNVDLAAIPNVKQIDGK 180
QY 181 YYYIGSDGPKKFNALTVNNKVLFPDKNTGALTDTSOYQFQKGLTKLNNDYTHNQIVNF 240
DB 181 YYYIGSDGPKKFNALTVNNKVLFPDKNTGALTDTSOYQFQKGLTKLNNDYTHNQIVNF 240
QY 241 ENTSLETIDNYVTADSWYRPKDILKNGKWTWASSSDLRPLLSMWDPKQTOIAYLNYN 300
DB 241 ENTSLETIDNYVTADSWYRPKDILKNGKWTWASSSDLRPLLSMWDPKQTOIAYLNYN 300
QY 301 QOGLGTGENYTADSSQESLNLAQTVQVKIETKISQTOQTQMLRDIINSFVKTPQNNWSQ 360
DB 301 QOGLGTGENYTADSSQESLNLAQTVQVKIETKISQTOQTQMLRDIINSFVKTPQNNWSQ 360
QY 361 TESDTSAGEKHLQCGALLYNSDKTAYANSYRLLNRTPTSTQTKPKYFEDNSSGGYDF 420
DB 361 TESDTSAGEKHLQCGALLYNSDKTAYANSYRLLNRTPTSTQTKPKYFEDNSSGGYDF 420
QY 421 LLANDIDNSNPVQAEQLNWLHLYLWNGYSIVANDPEANFDGVRVDAVDNVDNADLLQIASD 480
DB 421 LLANDIDNSNPVQAEQLNWLHLYLWNGYSIVANDPEANFDGVRVDAVDNVDNADLLQIASD 480
QY 481 YLKAHYGVDKSEKNAINHLSLEAWSNDNDPQNKDTKGAQLPIDNKLRLSLLYALTRPLE 540
DB 481 YLKAHYGVDKSEKNAINHLSLEAWSNDNDPQNKDTKGAQLPIDNKLRLSLLYALTRPLE 540
QY 541 KDAASKNEIRSGLEPVTNSLNRSAGKNSRMANYIPIRAHDSVQTVIAKIIKAQIN 600
DB 541 KDAASKNEIRSGLEPVTNSLNRSAGKNSRMANYIPIRAHDSVQTVIAKIIKAQIN 600
QY 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYYTOSNIPTAVALMSNKDSITRLYYGDMYS 660
DB 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYYTOSNIPTAVALMSNKDSITRLYYGDMYS 660
QY 661 DDGOYMATKSPYDAIDTLKAKIYAAGQDMKITYVEGDKSHMDWDYTGVLTSVRYGT 720
DB 661 DDGOYMATKSPYDAIDTLKAKIYAAGQDMKITYVEGDKSHMDWDYTGVLTSVRYGT 720
QY 721 GANEATDQGSSEATKTQGMVITSNPSLKLNDKVIYNMGAHKNQOEYRPLLLTKDGL 780
DB 721 GANEATDQGSSEATKTQGMVITSNPSLKLNDKVIYNMGAHKNQOEYRPLLLTKDGL 780


```
QY 1132 NNGHMYGLQOLNGEVOYFSLNGVQLRESFLENADSGKNYFGLHGNRYNSNGYYSFNDK 1191
DB 1101 NNGYMYTGAOSINGVNYFSLNGLRLDAILKNEGDGTAYYNGDRYENGYYQF--MSGV 1159
QY 1192 WRYFDASGVMAVGLKKTINGTQVFDODGYQVKGAWITGSDGKKRYFDDGSGMVAVRAN 1251
DB 1160 WRHFN--NGEMSUGLTVIDGQVYFDEMGVQAKGFVTTADGKIRYFDKSGMRYNRFTIE 1218
QY 1252 DRNGDWYLLNSDIALVGVQTINGKTYFYQDQKQIKGKIITD--NGKLYFLANSGLAR 1310
DB 1219 NEEGKWLXGDCGAAVTGSGTINGQHLVFRANGVQVKGFBVTDHGRISYDNGSGDQIR 1278
QY 1311 NIFATDSNNWYFSGDGVAVTGSQTIAGKLYFASDGKQVKGFSFTYTN--GKVHYTHADS 1369
DB 1279 NRPVRNAQGWYFDDNNGYAVTGARTINGQLLYFRANGVQVKGFEVTDYGRISYYDNGS 1338
QY 1370 GELQVRFEADKGNWYLLDSNGEALTGSORINDQVFFTRREGKQVKGVDVAYDERLLVY 1429
DB 1339 GQOIRNFRVNAQGWYFDDNNGYAVTGARTINGQHLVFRANGVQVKGFBVTDHGRISY 1398

RESULT 5
US-09-210-361-2
; Sequence 2, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 035/CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-210-361-2

Query Match 51.2%; Score 3817.5; DB 2; Length 1475;
Best Local Similarity 52.9%; Pred. No. 1.4e-221;
Matches 762; Conservative 230; Mismatches 395; Indels 53; Gaps 22;

QY 1 METKRYKMKHKVCHWTVAVASGLITLGTTLGSSVSASAEQOQTSKVVTKQSEDDKAA 60
DB 1 MDKRYKRLKRVKRWTVSVASAVWTL--TTLGGVLVRADSNESK-----QISDNTS 54
QY 61 SSSQTDAPKTKQAQTEQTAQSQAN--VADTSTSIKTPSQNIITTOANSDDKTVTNKS 119
DB 55 VVTANESNVITEATSQEAASSQTNHTVTTSSSTSVVNPKEVSNPVTUGETASNGEK 114
QY 120 EEAQTSERTKQSEAAQTASSQALTQAKAELTKORQTAQENKAPVDLAALPNVKQIDG 179
DB 115 LQNTQTTV--DKTSEAAANNISKOT--TEADTVDIDDSNAA-----NLQILEKLPNVKEIDG 167
QY 180 KYIYIGSDGQPKNFALTVNNKLVYFDKNTGALTDS--QYQFKQGLTKLNDYTHNQIV 238
DB 168 KYIYDNDGNKRVNFTLIADGKILHFD--TGAYTDTSIDTWNKDIIVTFRSNLYKKNQY 226
QY 239 NFENTSLETIDNYVTADSWYRPKDLKNGKWTWATSESDLRPLLSWNPDKQTQIAYLNY 298
```

```
DB 227 DRSAQFEHVDHYLTAEWYRPKYILKDGKWTQTSTKDFRPLLTMTWPDQETQRYVNY 286
QY 299 MNOQGLGTGENTYADSSQESLNLAOTVOVKIETKISQTOOTOWLRDINSFVKTOPNWN 358
DB 287 MNAQ--LGIINKTYDDTSNQLNLAATAIQAIEAKITTLKNTDMLRQTSIAFVKTSQAWN 345
QY 359 SOTESDTSAGEKHLGGGALLYSNSDK--TAYANSDYRLLRNRTPTSTQK--PKYFEDNSS 415
DB 346 SDSEKFPD---DHLQNGAVLYDNECKLTPYANSNYRILNRPTNTQTKDPRYTADNTI 401
QY 416 GGYDFLLANDIDNSNPVQABQLNWLHLYLWYSGSIIVANDPEANFDGVRVDVADVNVADLL 475
DB 402 GGYEFLANDVDNSNPVQAEQLNWLHFLWNEFNIIYANDPDANFDSIRVDAVDNVADLL 461
QY 476 QIASDYLLKHYGVDSKNAIINHLSITLBAWSNDPOYNKDTKGAQPIPKIKLRLSLLYAL 535
DB 462 QIAGDYLLKAAKGIHKNDKKAANDHLSILEAWSNDTPYLLHDDGDNMTNKNKLSLFLSL 521
QY 536 TRPLEKDSANKNEIRSGLEPVIITNSLNRSABGKNSERMANYIFIRAHDSVQVTIAKII 595
DB 522 AKPLNQ-----RSGMPLITNSLVNRTDDNAETAAPVPSYFIRAHDSVQDLADI 573
QY 596 KQAINPKTDLTFTLDELKQAFKIYNEDMRQAKKYTQSNIPYAYALMLSNKDSITRLYY 655
DB 574 KAEINPNVVGYSFTMEIEIKKAFIYNKOLLATEKKYTHYNTALSYALLLTNKSVPVRY 633
QY 656 GMYSDDGQYMATKSPYDAIDTLKARIKYAAGQDMKITVYEGDKSHMDWDYTGVLIS 715
DB 634 GDMFTDDGQYMAKHTINYEAIETLLKARIKYVGGQAMRNOQVGNSE-----IITS 684
QY 716 VRYGTGANEATDQGSSEATKQGMVITSNPSLKNQNDKVIYNMGAHKNQYRPLLLT 775
DB 685 VRYGKALKATDTGDTTFTSGVAVLEGNNPSLRLKASDRVVVNMGAHKNQYRPLLLT 744
QY 776 TKDGLTSTYSDAAAKSLYKNTKDELVPDASDIQYLNPNQVSGYLAVMVPVPGASQNDV 835
DB 745 TDNGIKAVHSDQEAAGLVRYTNDRGELIPTAADIKGANPQVSGYLVGVWVPVGA--LI 801
QY 836 RVAASNKANATQGV--YESSSALDSQLIYEGESNFQDFVTKDSDTNKKIAQNVLKFSW 893
DB 802 KMFALRLARPHQOMASVHQNAALDSRVMPFEGSFNFQAFATKKEEYVNVVIAKNVDFAEW 861
QY 894 GVTSPFEMAPQYVSSDGSFSDSIIONGAFEDRYDLAMSKNNKYGSOQDMINAVKALHS 953
DB 862 GYTFEMAPQYVSSDGSFSDSVIONGAFATDRYDLGISKPNKYGTADDLVKAKALHSK 921
QY 954 GIQVIADWPQIYNLPKGEVVTATRVNDYGBYRKDSSEIKNTLYAANTKNSGKDYQAKYG 1013
DB 922 GIKVMADWPQMYAPPEKEVVTATRVKYGTPVAGSQIKNTLYVVDGKSSGKQQAQYK 981
QY 1014 GAFISELAAPKPSIFNRTQISNGKKIDPSEKITANKAKYPNGTNTILGRGVGVYLDKNSD 1073
DB 982 GAFISELAAPKPSIFNRTQISNGKKIDPSEKITANKAKYPNGTNTILGRGVGVYLDKNSD 1041
QY 1074 KYFELKQNGO--TYLPKQMTNKEASTGFVNDGNGMTFYSTGYSQAKNSFVODAKGNWYFD 1131
DB 1042 TYFNISDNKEINFLPKTLNQDSQVGSYDGGYVYYSYSGYQAKNTFISEG--DKWYFD 1100
QY 1132 NNGHMYGLQOLNGEVOYFSLNGVQLRESFLENADSGKNYFGLHGNRYNSNGYYSFNDK 1191
DB 1101 NNGYMYTGAOSINGVNYFSLNGLRLDAILKNEGDGTAYYNGDRYENGYYQF--MSGV 1159
QY 1192 WRYFDASGVMAVGLKKTINGTQVFDODGYQVKGAWITGSDGKKRYFDDGSGMVAVRAN 1251
DB 1160 WRHFN--NGEMSUGLTVIDGQVYFDEMGVQAKGFVTTADGKIRYFDKSGMRYNRFTIE 1218
QY 1252 DRNGDWYLLNSDIALVGVQTINGKTYFYQDQKQIKGKIITD--NGKLYFLANSGLAR 1310
DB 1219 NEEGKWLXGDCGAAVTGSGTINGQHLVFRANGVQVKGFEVTDHGRISYDNGSGDQIR 1278
QY 1311 NIFATDSNNWYFSGDGVAVTGSQTIAGKLYFASDGKQVKGFSFTYTN--GKVHYTHADS 1369
```

Db 1279 NRVFNAQOQWYFEDNNGYAVTGARTINGQLLYFRANGVQVKGFEVTDYGRISYYDGN 1338
QY 1370 GELQVRPEADKQGNWYLDNGBEALTSQORINDQVFFTRGKQVKGVDVADERLLVY 1429
Db 1339 GQOIRNFRVNAQOQWYFEDNNGYAVTGARTINGQLLYFRANGVQVKGFEVTDYGRISY 1398

RESULT 6

US-09-740-274-2
; Sequence 2, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PR1
; ORGANISM: Streptococcus mutans
US-09-740-274-2

Query Match 51.2%; Score 3817.5; DB 2; Length 1475;

Best Local Similarity 52.9%; Pred. No. 1.4e-221;

Matches 762; Conservative 230; Mismatches 395; Indels 53; Gaps 22;

QY 1 METKRYKHKKHVVAVASGLITGLTGLSSVSAETEQQTSKVVTKQSEDDKAA 60
Db 1 MDKRYKLRKKRWVTVSVASVMTL-TTLSSGLVKADSNESK-----QISNDSNTS 54
QY 61 SSSOTDAPKTQAOEQTOAQOAN-VADTSTSTTKETPSONITTOANSDDKTVTNTKS 119
Db 55 VTANESNVITEATSKQAASSQTNHTVTSSSTSVVNPKEVSNPYTVGETASGEK 114
QY 120 EBAQTSERTKQSEEAQTASSQALTOAKAELTKQRTAAQENKNPVDLAAIPNVKQIDG 179
Db 115 LQNTTV-DKTSEAAANNISQOT-TEADTDVIDDSNAA-----NLQILEKLPNVKEIDG 167
QY 180 KYITGSDQPKNFALTVNNKVLFDKNTGALTDS-QYQKQGLTKLNDYTHNQIV 238
Db 168 KYITDNNKQVNTFTLADGKILHFE-TGAYTDTSDTINKDIIVTTRSNLYKKYNQY 226
QY 239 NFENTSLETIDNVVADSVRPKDILKNGKVTASSEDLRLPLLSMWDPDKOTQIAYLNY 298
Db 227 DSAQSFEDHVLTHAESWYRKYILKQKQVTKQTEKDFRLLMTWEPDQETQRYNY 286
QY 299 MNQQLGIGENTYADSSQSLAAQTVQVKIETKISQQTQWLRIIINSFVKTPQNNW 358
Db 287 MNAQ-LGINKTYDDTSNQLQNLIAAATIQAIEAKITTLKNTDWRQTISAFVKIQSAMN 345
QY 359 SQTESDTSAGEKHLOGGALLVNSDK-TAYANSYRLNRPRTSQTK--PKYFEDNS 415
Db 346 SDSEKFPD----DHLQNGAVLYDNEGKLTTPYANSYRLNRPRTNQTKGKDPRTADNTI 401
QY 416 GGYDFLLANDIDNSNPVQAEOLNHLFNMFGNIYANDPDANPDSIRVDAVDNVDADLL 475

RESULT 7

US-08-793-824-2
; Sequence 2, Application US/08793824
; Patent No. 5981838
; GENERAL INFORMATION:
; APPLICANT: Simpson, Christine Lynn

Db 402 GGYEFLLANDVDSNPVQAEOLNHLFNMFGNIYANDPDANPDSIRVDAVDNVDADLL 461
QY 476 QIASDYLKAHVGVDSKSEKNAIHLISILEAWSNDNDPOYKNDTKGAOLPIDNKLRLSLLYAL 535
Db 462 QIAGDYLKAAGIHKNDKAANDHLISILEAWSNDNDPYLHDDGDNNMINMDNKLRLSLFLSL 521
QY 536 TRPEKDSASKNKNEIRSGLEPVITNSLNNRSGKSEKSEKSEKSEKSEKSEKSEKSEKSEK 595
Db 522 AKPLNQ-----RSGNPLITNSLVNRTDNDNAETAAPVSFIRAHDSVQDLADI 573
QY 596 KQAINPKTDGLTFTLDELKQAFKIYNEBDMQAKKYYTQSNIPATYALMLSNKDSITRLYY 655
Db 574 KAEINPNVVGYSFTWEEIKKAFIYNKDLALATEKKYTHYNTALSAYALLTNKSSVPRVY 633
QY 656 GMYSDDCQYMATKSPPYDAIDTLKARIKYAAGQDMKIITYVEGDKSHMDWDYTGVLTS 715
Db 634 GDMFTDDGQYMAHKTINYEAIETLLKARIKYVSGQAMRNQOVGNSE-----IITS 684
QY 716 VRYGTGANEATDOGSEATKTQGMVITSNPSLKLNDKVIYNNMGAHKNQYRPLLT 775
Db 685 VRYGKALKATDGTDRTRTSGVAVIEGNNPSLRUKASDRVVVNNMGAHKNQYRPLLT 744
QY 776 TKDGLTSYTSDAAKSLYKRTNDKGLVFDASDIQGYLNPQVSGYLAVWVPVGSADNDV 835
Db 745 TDNGIKAYHSQEAAGLVRYTNDRCGLIFTAADIKGYANPQVSGYLVWVPVGA--LI 801
QY 836 RVAANKANATQV--YESSSALDSQLIYEGFSNFQDFVTKDSDTNKKIAQNVOLFQSW 893
Db 802 KMFALRLARPHQOMASVHQNAALDSRVMEFEGSNFQAFATKKEBYTNVVIKKNVDFAEW 861
QY 894 GVTSPFEMAPQVYSSSEDSGLSDSIIONGAFEDRYDLAMSNNKYQSGQODMINAVKALHS 953
Db 862 GVTDFEMAPQVYSSDGSGLSDSIIONGAFEDRYDLAMSNNKYQSGQODMINAVKALHS 921
QY 954 GIQVIADWVPDQIYNLPQKEVVTATRVNDYGEYRDKSEIKNTLYAANTSKNGKDYQAKYG 1013
Db 922 GIKVMADWVPDQMYAFPEKEVVTATRVNDYGEYRDKSEIKNTLYAANTSKNGKDYQAKYG 981
QY 1014 GAFLELAAPPSIFNRTQISNGKKIDSEKIKTANKAKYFNGTNIILGRGVGVVLKDNASD 1073
Db 982 GAFLELOAKYPPELPARKQISTGVPMPSVKIKOWSAKYFNGTNIILGRGVGVVLKDOATN 1041
QY 1074 KYFELKGNQ--TYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVODAKGNWYFD 1131
Db 1042 TYFNISDNKEINFLPRTLLNQDSQVGFSDGKGYVYTSYGQAKNTFISEG-DKWWYFD 1100
QY 1132 NNGHMYVGLQNLNGEVQYFLSNGVQLRESFLENADGSKNYFGLGNRYSNGYYSFNDNSK 1191
Db 1101 NNGYVMTGAQSGINGVNYFLSNGVQLRESFLENADGSKNYFGLGNRYSNGYYSFNDNSK 1159
QY 1192 WRYFDASGVMAVGLKTINGTOYFDQDGYOVKGAITGSDGKKRYPDQSGNMAVNRPAN 1251
Db 1160 WRHFN-NGEMSGLTVIDQVQYFDEMGMGYQAKGKFTVADGKIRYFDPKOSGNMYNRFT 1218
QY 1252 DKNGDNYLNSDIALVGVQTINGKYFYFGDQKQIKGKIITD-NGKLYFYLANSGELAR 1310
Db 1219 NEEGKWLIGEDAAVTGSQTINGQHLIFRANGVQVKGFEVTDYHGRISYDGNSDQIR 1278
QY 1311 NIFATDSQNNWYFSGDGVAVTGSQTIAGKLYFASDQKQVKGSEKSEKSEKSEKSEKSEK 1369
Db 1279 NRVFNAQOQWYFEDNNGYAVTGARTINGQLLYFRANGVQVKGFEVTDYGRISYYDGN 1338
QY 1370 GELQVRPEADKQGNWYLDNGBEALTSQORINDQVFFTRGKQVKGVDVADERLLVY 1429
Db 1339 GQOIRNFRVNAQOQWYFEDNNGYAVTGARTINGQLLYFRANGVQVKGFEVTDYHGRISY 1398

QY 1400 RINDQVFTTREGKQVKGVDVAY 1421
DB 1537 VINGKQLYFDGSGRQVKGRIYV 1558

RESULT 8

US-09-210-361-4
; Sequence 4, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE REFERENCE: Latexes in Paper Manufacture
; CURRENT APPLICATION NUMBER: US/09/210,361
; EARLIER FILING DATE: 1998-12-11
; EARLIER FILING DATE: 1998-01-16
; EARLIER FILING DATE: 1995-06-07
; EARLIER FILING DATE: 1998-01-20
; EARLIER FILING DATE: 1995-06-07
; EARLIER FILING DATE: 1995-06-07
; EARLIER FILING DATE: 1998-01-16
; EARLIER FILING DATE: 1998-01-16
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-210-361-4

Query Match 48.1%; Score 3591.5; DB 2; Length 1375;
Best Local Similarity 51.9%; Pred. No. 5.5e-208;
Matches 721; Conservative 215; Mismatches 350; Indels 103; Gaps 22;

QY 1 METKRYKMKVKKCHWTVAVASGLITLCTTLTGSSVSASAEQTSQSDKVVTKQSDKAA 60
DB 1 MEKKVRFKVRKRVKRVTVSIAVAVTL-TSUGSLVKADS-----TDRQQA 47
QY 61 SSSQTDAPKTQAOEQQAQANVADTSITKETPSQNIITQANSDDKTVTNTKSE 120
DB 48 VTESQASLVTTSEA-AKETLTATDTSTATSQPTATVTDNVSTNQ-----TNTTAN 101
QY 121 EAQTSERTKQSEEAQTASSQALTOAKA--ELT----- 152
DB 102 TANFVVKPTTTEQAQNTSDSKIIITTSKAVNRLTATGKFPVANNNTAHPKTVTDKIVPIK 161
QY 153 -----KQRTAAQENKPNVDLAAIPNVK---QIDGKYIIGSDGPKQKFNALTWNKYL 203
DB 162 PKIGKLKQSSLSQD-----DIAALGNVKNIRKVKYKYYKEDGLQKNYALNINGKTF 216
QY 204 YFDKNTGALTDSYQFKQGLTKLND-----YTPHQIIVNFENTSLETIDNVVTADSWYR 259
DB 217 FFDE-TGALSNNLTSPSKGNIT--NNDNTNSFAQYNQVYSTDVANFEHVDHILTAESWYR 273
QY 260 PKDILKNGKWTWASSSDRLPLMSWPKQTOIAVLNMNOGLGTGENYTDSSQBSL 319
DB 274 PKYILKDGKWTQSTEKDFRPLMTWPDQETQRQYVNTMNAQ-LGIHQTYNTATSPLOL 332
QY 320 NLAAQTVQVKIETKISQTOQTQWLTDIINSFVKTPQPNWNSQTESDTSAGEKHLOGGALL 379
DB 333 NLAAQTIQTKIBEKITAENKTNWLRQTIISAFVKTOQSAWNSDSEKPPD-----DHLQKGALL 388
QY 380 YSNSDK-TAYANSYDLLNRTPTSTQTK--PKYFEDNSSGGYDFLLANDIDNSNPVQAE 436
DB 389 YSNSKLTQANSYRILNRTPTNQTKGDPRTADRTIGGYEFLANDVDNSNPVQAE 448

QY 437 QLNWLHLYMNYGSIIVANDPEANFDCVRVDAVDNVDNADLLOIASDYLKAHYGVYDSEKNAI 496
DB 449 QLNWLHFLMNFNYANDPNDANFDSIRVDADVNDVADLLOIAGDYLKAAGHKNKADKAN 508
QY 497 NHLISLEAWSNDPOYNKDTKGAQLPIDNKLRLSLLYALTRPLEKASNKIRSGLEPV 556
DB 509 DHLISLEAWSYNDTPYLHDDGDNMNMNRLSLSLYLAKPLNQ-----RSGMNP 560
QY 557 ITNSLNNRSAGKNSERMANYIFIRAHDSVQTVIAKIIKAQINPKTDGLTFTLDELKQA 616
DB 561 ITNSLVNRTDDNAETAAPSYSFIRAHDSVQDLIRNIIRTEINPNVVGYSFTTEIKKA 620
QY 617 FKYNEDMRQAKKKTQSNIPATYALMNSKDSITELYYGDMYSDGQVATKSPYDAI 676
DB 621 FEIYNKOLLATEKKYTHYNTALSALLTNKSSVPRVYVYDGMPTDDGQVMAKHTINYEAI 680
QY 677 DTLKARIKYAAGQDMKITTYVEGDKSHMDWDYTGVLTSVRYGTGANEATDQSEATKTK 736
DB 681 ETLKARIKYVSGGQAMRNQOVGNSE-----IITSVRYGKALKATDGTDRTRTS 731
QY 737 GMAVITSNPSLKNQNDKIVNMGAHKNQBYRPLLLTTKDLGTSYTSDAAKSLYRKT 796
DB 732 GVAVIEGNPNSLRLKASDRVVNMGAHKNQBYRPLLLTTDNGIKAYHSDQBAAGLVRYT 791
QY 797 NDKGELVFDASDIQCLNPOVSGYLAVVWPGASDNQDVRVAASNKANATQVYESSAL 856
DB 792 NDRGELIFTAADIKGYANPQVSGYLGWVPVGAADQDVRVAASTAPSDTGKSVHQAAL 851
QY 857 DSOLIVEGFSNFDQFVTKDSYTNKKIAQNVOLFWSGWVTSFEMAPQYVSSDGSFLDSI 916
DB 852 DSRVMEFEGSNFQAFATKEEYTNVVIKNDVKFAEWGTVDFEMAPQYVSSDGSFLDSV 911
QY 917 IQNGYAFEDRYDLAMSKNKYQSDQDMINAVKALHKSQIVTADWVVDQIYNLPKGEVVT 976
DB 912 IQNGYAFEDRYDLGHSKPKYGTADLVKAIKALHSKGIKVMADWVVDQMYALPEKEVVT 971
QY 977 ATRVNDYGEYRKDSIKNTLYAANTKSNKGDKYQAKYGGAFSLSELAAKYPSINRTQISNG 1036
DB 972 ATRVDKYGTPVAGSQIKNTLYVVDGKSGKQQAQYGGAFBELOAKYPELFARKQISTG 1031
QY 1037 KKIDSEKITAMKAKYFNGTNILGRGVYLVKDNASDKYFELKGNQYLPKQMTN----- 1091
DB 1032 VPMDFSVKLIKQWAKYFNGTNILGRGAGYLVKQATNTYFSLVSDNTFLPKSLVNPNGHT 1091
QY 1092 KEASTGFVNDGNGMTFTYSTGYQAKNSFVQADAKGNNYFDDNNGHMYVGLQQLNGEVQYFL 1151
DB 1092 SSSVTLGLVFDGKGYVYVYSTSGNQAKNAFI-SLGNNNYFDDNNGYVMTGASINGANYFL 1150
QY 1152 SNGVOLRESFLENADGSKNYFGLGNRYSGYYSFDNDSKWEYFDASGVMAVGLKTINGN 1211
DB 1151 SNGIQLRNAIYDNGNKKVLSYCGNDGRRYENGYYLF--GQWRYFO-NGIMAVGLTRVHGA 1207
QY 1212 TOYFDDQGVQVKGAMITGSDGKKRYFDDGSGNMVNRFANDKNGDWYLLNSDGIALVGVQ 1271
DB 1208 VOYFDASGEQAKQPIITADGKLRVDRDSGNQISNRFVNSKGEWFLFDHNGVATGTV 1267
QY 1272 TINGKTYFQDGQKQIKGIITD-NGKLYFLANGELARNIPATDSQNNWYFSGDGA 1330
DB 1268 TENGRLYFKPNGVQAKGFIIRDANGYLRYYPNSGNEVRNFRVNSKGEWFLFDHNGIA 1327
QY 1331 VTGSQTIAG 1339
DB 1328 VTGARVWNG 1336

RESULT 9
US-09-740-274-4
; Sequence 4, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD

US-09-604-957-4

Query Match 37.6%; Score 2808; DB 2; Length 545;
Best Local Similarity 100.0%; Pred. No. 2.6e-161;
Matches 545; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 421 LLANDIDNSNPVQAEQLNLHLHLYLMNYGSIIVANDPEANFDGVRVDAVNNDLQIAD 480
DB 1 LLANDIDNSNPVQAEQLNLHLHLYLMNYGSIIVANDPEANFDGVRVDAVNNDLQIAD 60
QY 481 YLKAHYGVDKSEKNAIHLHSILLEAWSNDNDPQNKDTKGAQLPIDNKLRLSLLYALTRPLE 540
DB 61 YLKAHYGVDKSEKNAIHLHSILLEAWSNDNDPQNKDTKGAQLPIDNKLRLSLLYALTRPLE 120
QY 541 KDAANKNEIRSGLEPVITNSLNNRSGKNSERMANYIFIRAHDSVQTVIAKIIKAQIN 600
DB 121 KDAANKNEIRSGLEPVITNSLNNRSGKNSERMANYIFIRAHDSVQTVIAKIIKAQIN 180
QY 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYYTQSNIPATAYALMLSNKDSITRLYYGDMYS 660
DB 181 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYYTQSNIPATAYALMLSNKDSITRLYYGDMYS 240
QY 661 DDGQYMATKSPYYDAIDTLKARIKYAAGQDMKITYVEGDKSHMDWDYTGVLTSVRYGT 720
DB 241 DDGQYMATKSPYYDAIDTLKARIKYAAGQDMKITYVEGDKSHMDWDYTGVLTSVRYGT 300
QY 721 GANEATDQSEATKTQGMAVITSNPNLSKLNQNDKVIYVNMGAHKNQYRPLLLTTKQGL 780
DB 301 GANEATDQSEATKTQGMAVITSNPNLSKLNQNDKVIYVNMGAHKNQYRPLLLTTKQGL 360
QY 781 TSYSDDAAKSLYRKTNDKGLVFDASDIQGYLNPQVSGYLAWVPVPGASDNQDVRVAAS 840
DB 361 TSYSDDAAKSLYRKTNDKGLVFDASDIQGYLNPQVSGYLAWVPVPGASDNQDVRVAAS 420
QY 841 NKANATGOVYESSALDSQLIYEGFSNFODFTVKDSYTNKKIAQNVQLFKSWGVTSPFM 900
DB 421 NKANATGOVYESSALDSQLIYEGFSNFODFTVKDSYTNKKIAQNVQLFKSWGVTSPFM 480
QY 901 APQYVSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSQQDMINAVKALHKSIGQVIAD 960
DB 481 APQYVSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSQQDMINAVKALHKSIGQVIAD 540
QY 961 WVPDQ 965
DB 541 WVPDQ 545

RESULT 11
US-09-995-749A-10
; Sequence 10, Application US/09995749A
; Patent No. 6867026
; GENERAL INFORMATION:
; APPLICANT: VAN GEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: BO43388-CIP
; CURRENT APPLICATION NUMBER: US/09/995, 749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-995-749A-10

Query Match 37.3%; Score 2780; DB 2; Length 545;

Best Local Similarity 99.6%; Pred. No. 1.3e-159;
Matches 544; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 421 LLANDIDNSNPVQAEQLNLHLHLYLMNYGSIIVANDPEANFDGVRVDAVNNDLQIAD 480
DB 1 LLANDIDNSNPVQAEQLNLHLHLYLMNYGSIIVANDPEANFDGVRVDAVNNDLQIAD 60
QY 481 YLKAHYGVDKSEKNAIHLHSILLEAWSNDNDPQNKDTKGAQLPIDNKLRLSLLYALTRPLE 540
DB 61 YLKAHYGVDKSEKNAIHLHSILLEAWSNDNDPQNKDTKGAQLPIDNKLRLSLLYALTRPLE 120
QY 541 KDAANKNEIRSGLEPVITNSLNNRSGKNSERMANYIFIRAHDSVQTVIAKIIKAQIN 600
DB 121 KDAANKNEIRSGLEPVITNSLNNRSGKNSERMANYIFIRAHDSVQTVIAKIIKAQIN 180
QY 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYYTQSNIPATAYALMLSNKDSITRLYYGDMYS 660
DB 181 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYYTQSNIPATAYALMLSNKDSITRLYYGDMYS 240
QY 661 DDGQYMATKSPYYDAIDTLKARIKYAAGQDMKITYVEGDKSHMDWDYTGVLTSVRYGT 720
DB 241 DDGQYMATKSPYYDAIDTLKARIKYAAGQDMKITYVEGDKSHMDWDYTGVLTSVRYGT 300
QY 721 GANEATDQSEATKTQGMAVITSNPNLSKLNQNDKVIYVNMGAHKNQYRPLLLTTKQGL 780
DB 301 GANEATDQSEATKTQGMAVITSNPNLSKLNQNDKVIYVNMGAHKNQYRPLLLTTKQGL 360
QY 781 TSYSDDAAKSLYRKTNDKGLVFDASDIQGYLNPQVSGYLAWVPVPGASDNQDVRVAAS 839
DB 361 TSYSDDAAKSLYRKTNDKGLVFDASDIQGYLNPQVSGYLAWVPVPGASDNQDVRVAAS 419
QY 840 NKANATGOVYESSALDSQLIYEGFSNFODFTVKDSYTNKKIAQNVQLFKSWGVTSPSE 899
DB 420 NKANATGOVYESSALDSQLIYEGFSNFODFTVKDSYTNKKIAQNVQLFKSWGVTSPSE 479
QY 900 MAPQVYSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSQQDMINAVKALHKSIGQVIA 959
DB 480 MAPQVYSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSQQDMINAVKALHKSIGQVIA 539
QY 960 DWVPDQ 965
DB 540 DWVPDQ 545

RESULT 12
US-09-499-203-2
; Sequence 2, Application US/09499203
; Patent No. 6570065
; GENERAL INFORMATION:
; APPLICANT: KOSSMANN, Jens
; APPLICANT: WELSH, Thomas
; APPLICANT: QUANZ, Martin
; APPLICANT: KNUTH, Karola
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternansucrase
; FILE REFERENCE: 147-196P
; CURRENT APPLICATION NUMBER: US/09/499,203
; CURRENT FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2057
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-09-499-203-2

Query Match 36.2%; Score 2699.5; DB 2; Length 2057;
Best Local Similarity 38.1%; Pred. No. 6.2e-154;
Matches 651; Conservative 221; Mismatches 493; Indels 343; Gaps 45;

QY 2 ETKRRYKMHKVKHVVAVASGLITLGTTLTGSSVSAETQOTSDKVV----TQKSEDD 57
DB 5 ETVTRKKLYKSGKVMVAATAFAVLGVSTVT---TVHADTNSNVAVKQINNTGTNDGSEK 61

QY 58 KAASQSOTDA-----PKTKQ--AQTEQTAQOANVA-----DTS--- 91
Db 62 KVPVSTNDSLKQGTGDFWSDGNRVQKTNQIILLTABQLKXNNEKNSLVISDDTSK 121
QY 92 --TSITKTPSQ-----ITTOANSDDKT----- 113
Db 122 DDENISKQTKIANQOTVDTAKGLTTLNSLDPITGHHYENHNGYFVVIDASGKQVTLQNI 181
QY 114 -----VNTKSEEAQTSERTQSB 133
Db 182 DGNLQYFDDNGYQVKGSPRVNGKHIFYSVTGKASSNVDIVNGKAQGYDAQGNQLKXSV 241
QY 134 EAQTTASS-----QALTO----- 146
Db 242 VADSSQTYFYFGNGOPLTGLQITDGNLQYFNGQGVQIKGGFQDVNNKRIYFAPNTGNV 301
QY 147 AKAEALTQKQTAQAENKPV-----DLAA-----IPNVKQIDGKYYIGSDGO 189
Db 302 ANTEILINGKLQDRDANGQVKNAFSKDVAGNTFYDPANGVMLTGLQITSGKTYLDEQGH 361
QY 190 PKKNFALTNNKVLFPDKNTGALTDTDSYQFQKGLFKLANDYTPHNNQIVNFENTSLETTID 249
Db 362 LRKNYAGTFNNOFMVFPDADTGAGTAIEYQFQGLVQSQSNENTPHNAAKSYDKSSPENVD 421
QY 250 NYVTADSWYRPKDIILKNGKWTWTASSESDLRPLLMSWPKQTOIAYLNNVMOOGLGTGEN 309
Db 422 GYLTDATWYRPDIILKNGGDTWTASTETDWRPMLTWWPDKQYQANYLNFMSKGLGITT 481
QY 310 YTADSSQESNLAAQTVQVKIETKISOTQOTOWLRDIINSFVKTOPNWSQTESDTSAGE 369
Db 482 YTAATSQKTLNDAAFIQTAIEQOISLKSTEWLRDAIDSFVKTOANNKQTEDEAFDG- 540
QY 370 KDHLOQGALLY--SNSDKTAYANS--PYRLNRTPTSTQTKPKFYFEDNSSG--GYDFLLANDI 426
Db 541 LOWLQGGFLAYQDDSHRTPNTDSGNRKLGRQPINIGS-----KOTTDGKSGSEFLANDI 596
QY 427 DNSNPVQAEQLNWLHYLMMYSGIIVANDPEANFDGVRDAVNADLLOIADSYLKAHY 486
Db 597 DNSNPVQAEQLNWLHYLMMYSGITGNNDNANFDGIRDAVNDVADLLKIAGDYFKALY 656
QY 487 GVDKSEKAINHLSILEAWSDDNDPOVNDTKGAQLPID-----NKL- 528
Db 657 GTDKSDANAKHLSILEWNGKDPQVNVQNGNAQTMDYTVTSQFGNSLTHGANNRNNW 716
QY 529 --LSLLYALTREPLEKSDASNKNEIRSGLEPVIITNSLNRSAEKGNSRMANYIFIRAHDS 586
Db 717 YFLDTGYYLNGDLNKKIVDKNRPNSS-----TLVNRANSQDTKVPINYSFVRAHYD 769
QY 587 VQTUTAK-----IIKAQINPKYDGLTTLDELKQAFKIYNEDMQAK--KKYTSQNTPT 638
Db 770 AQDPIRKAMIDHGIKNNMQD-----TFTFDQAQGMFEYKQDQENPSPGFKKYNDYNLPS 823
QY 639 AYALMLSNKDSITRLYYGDMYSDDOYMATKSPYYDAITLLKARIKYAAGQDMKIYV 698
Db 824 AYAMLLTNKQTVPRVYGDMLYLEGQYMEKGYIYNPVSALLKARIKYVSGSQWTAT--- 880
QY 699 EGDKSHMDW--DYTGVLTSVRXYGTGA---NEATDQSGEATKTQGMVITSNPCLKNQ 752
Db 881 --DSSGKDLKQGETDLLTSVRFKGMTSDQTTTQDQNSQDYKXQGIQVIVGNPDLKLN 938
QY 753 NDKVIVNNGAAHQKQYRPLLLTTTKDGLTSYTSDAAAKSLYKRTKNDKGELVFDAS----- 807
Db 939 DKTITLHMGKAHQKLYRALVLSNDGSDIVYDSDDKAPT--RTNDNGDLIFHKNTFKV 996
QY 808 -----DIOGLYNPQSGVLAWVVPVGSADNQVR--VAASNKANATQVYVSSSALDS 858
Db 997 QDGTIINTEMKGSNLALISGLVGVWVPVGSADSDQARTVATESSSSSNDGSVFHSNAALDS 1056
QY 859 QLIYEGFNFQDFVTKDSDYTNKKIAQNVQLFKSMGVTSFEMAPQYVSSSEDG-----SFL 913
Db 1057 NVIYEGFSNFQAMPTSPQSTNVVATKANLFLKELGITSFELAPQYRSSGDTNYGGMFL 1116
QY 914 DSIITQNGYAFEDRYDLAMSKNN-----KYGSDQDMINAVKALHKSQIVTADWVPDQIY 967

Db 1117 DGLFNNGYAFTDRYDLGFNKAQDGNPNPTKYGTDDLRNAIEALHKNQMOAIADWVPDQIY 1176
QY 968 NLPKGKVVVATRVNDYGEYRKDSEIKNTLYAAANTKSGKDYQAKYKGAPLSELAACKYPSI 1027
Db 1177 ALFPGKVVVATRVDERGNQLKOTDFVNLIVYANTKSSGVDYQAKYGEFLDKUREEYPSL 1236
QY 1028 FNRTQISNGKKIDPSEKITAUKAKYFNGTNIILGRGVYVVLKONASDKYFEL--KGNQTYLP 1086
Db 1237 FKQNOVSTGPIDASTKIKQWSAKYMGNTNIIHRGAYVYVLKOWATNQYFNIAKTNEVFLP 1296
QY 1087 KQMTNKEASTGFPVNDGNGMTFYSTSYQAKNSFVQDAKGNWYFNDNGHMYGLQ----- 1141
Db 1297 LQLOKNDQAGTFISDASGVKYYSISGYQAKDTPIEDGNGNWWYFDDGYMVRSSQGSNPI 1356
QY 1142 -----QLNGEVOYFLSNGVQLRESFLENADGSKNYFCHLGNRRYSNGYVSFDDNSKWR 1193
Db 1357 RIVETSVNTRNGN--YFPMFNGVELKRGFGTDSNGVYFDDQCKMVRDKYIN--DDANNFY 1414
QY 1194 YFDASGVMAVGLKTNGNT--QYFDQDGYQVKGAWITGSDGKKRYFDDGSGNMAVNRFAN- 1251
Db 1415 HLNVDGTMRSRGLFKFDSDTLQYFASNGVQIKDSYAKDSKGNKYYPDSATGNNDTGKAQTV 1474
QY 1252 DKNGDWYLYNSOGIALVGQV-----IN 1274
Db 1475 DNGYIYITIDSANNTIGVNTDYTAYITSSLRDGLFANAPYGVVTKDQNGNDLKWQYIN 1534
QY 1275 GXTYIFGO-----DGKQIKGIITDNG--KLKYFLANSSELARINPATDSQNNW- 1321
Db 1535 HTKQYEGQOVQVTRQYVTSKGVSWNLITPAGDGLQGRULWVDSRALTMPTPKTMNQISFI 1594
QY 1322 --YFGSDGVAVTGSQTIAGKKLYFASDGKQVKGSFVYTKGVHYHADSGELQVNRFEAD 1380
Db 1595 SYANRNDGLFLNAPYQVKGQYL--AGMSNQYKQGVTIAGVANVSGKOWSLISFN----- 1647
QY 1381 KQGNWYLYDSNGEAL--TGSQRINDQRFV 1407
Db 1648 --GTQYWIDS--QALNTNFTTHDMNQKF 1671
RESULT 13
US-09-604-957-3
; Sequence 3, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; PRIOR FILING DATE: 2000-08-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1278
; TYPE: PRT
; ORGANISM: Lactobacillus reuteri
US-09-604-957-3
Query Match 32.8%; Score 2446; DB 2; Length 1278;
Best Local Similarity 46.7%; Pred. No. 6.1e-139;
Matches 508; Conservative 170; Mismatches 310; Indels 100; Gaps 25;
QY 174 VKQIDGKYYIIG-SDGQPKKNFALTNNKVLVFDKNTGALTDTDSYQFQKGLTKLNNDYT 232
Db 235 LQTINGQYIYDPTTGQPKRKNFLQSGNNWYIFDSDTGVGTNVALELOQAKGVTSNEQYR 294
QY 233 PHNOIVNFPENTSLETTIDNYVTADSWYRPKDIILKNGKWTWTASSESDLRPLLMSWFPDQIY 292

```
Db 295 NGNAAYSDKSIENNVGYLTDWYRPKQILKOGTTWTDTSKETDMRPILMWNTLTQ 354
QY 293 IAYLNYMNOOGLGTGEN-----YTADSSOESLNLAQTVQVKIETKISOTOQTOWLR 344
Db 355 AYLNYMNOHG-----NLLPSALPFNADADPAELNHYSEIVQONIEKRISETGTMDLR 409
QY 345 DIINSFVKTPQPNWNSQTSRDSAGSKDHLQGGALLYSDKTYANSDYRLLNRPPTSOT 404
Db 410 TLMHDFVTNNPMWNKDSNVNFSGIQ--FQGFLLKXENSDLTPYANSRYRLGRPIN-- 465
QY 405 GKPKYFEDNSGGYDFLLANDIDNSNPVVQABQLNWLHYLMNYGSI VANDPEANFDGVRV 464
Db 466 -----IKDQYRGQBFLLANDIDNSNPVVQABQLNWLHYLLNFGTITANNDOANFDSVRV 520
QY 465 DAVDNVNADLLOIASDYLKAHYGVDKSEKNAIHLISLEAWSDNDPOYNKDTKGALPID 524
Db 521 DAPDNIDADLMNIAQDYFNAAYGMD-SDAVSNKHINILEDWNHADPEYFNKIGNQLTMD 579
QY 525 NKRLSLLYALTRPLEKASNKNEIRSGLEPVITNSLNNRSAGKNSERMANYIFIRAH 584
Db 580 DTIKNSLNHGLS-----DATN----RWGLDAIVHQSADRENNTENNVVIPNYSFVRAD 630
QY 585 SEVQTVIAKIIKAQINPKTDGLTFTLDELQAFKIYNEDMRQAKKYTOSNIPTAVALML 644
Db 631 NNSQDOIQNAIR-DVTGK-DYHTFTFEDEQKIDAYIQDQNSTVKKNLYNPASVAILL 688
QY 645 SNKDSITRLYYGDMYSDGOYNATKSPYYDAIDTLKARIKYAAGQDMKIITYVEGDKSH 704
Db 689 TNKDITPRVYGDLYTDCGOYMEHQTRYDDTLNLLKSEVKYVAGGOSQMTSVGNNN- 747
QY 705 MDWDVTGLVTSVRYGTGANEATDQGEATKTQGMAVITSNNSPLKLNQNDKVI VNMGAH 764
Db 748 -----ILTSVRYGKGAMTATDGTDETRTQIGVGVVSNTPNLKLGVDNKVVLHMGAAH 800
QY 765 KNOEYRPLLTTKDGLTSTSDAAKSLYRKNTDKELVFDASD-----IQGY 812
Db 801 KNQOYRAAVLTTDGVINYSDOGAP--VAMTDENGDLVLSHNLVNGKEEADTAVQGY 858
QY 813 LNPQVSYLAVWPVYGASDNQDVRVAASNKANATQVYESSALDSOLITYEGFSNFQDFV 872
Db 859 ANPDVSYLAVWPVYGASDNQDARTAPSTEKNSGNSAYETNAAFDSNVIFAFSNVYTP 918
QY 873 TKDSYTNKKIAQNVOLFSGWGTGFEMAPQVSVSDEGSLDSIIQNGVAFEDRYDLAMS 932
Db 919 TKESERANVRIAQNADFFASLGTFSFEMAPQVNSKKDRFTLDSITDNGVAFTRYDYLGS 978
QY 933 KKKYGSQODMINAVKALHKSIGIOVADWVPDQIYNLPKGEVVVTRVNDYGEYKDSI 992
Db 979 EPNKYGTDEDLRNAIQALHKLQVMAHWVPDQIYNLPKGEVATVTRVDDRGVWKRDAII 1038
QY 993 KNTLYAANTKSGKDYQAKYGAFISELAAKYPSIFNRTOISNGKKIDPSEKITAWKAKY 1052
Db 1039 NNNLYVWNTIGGG-EYQKYGAFDLKLOKLYPEFTKKQVSTGVAIDPSQKITWSAKY 1097
QY 1053 FNGTNILGRGVGVYLVKONASDKYFELKGNQOT--YLPKQMT--NKEASTGFV--NDNGMWT 1106
Db 1098 FNGTNILHRSGVLKADGG-QYVNL-GTTTQOFLPIQLTGKKGQNEGFAVGNDGN-YY 1154
QY 1107 FYSTSGYQAKNSFVQDAKKNYYFDNNGHMYVGLQOLN-----GE--VQYFSLNGVLR 1160
Db 1155 FYDLAGNMVKNFTFIEDSVGNWFFFDQDGHQVKNKHFPVDVDSYGEKGTFFLKNKGSFRRG 1214
QY 1161 FLENADGSKNYFCHLGNRYSGNYSFNDNSKWRYPDASGVMAVGLKTNGNTQYF--DQD 1218
Db 1215 LVQTF-----DNGTYFDNYGK-----MVRNQTINAGAMIYTLDEN 1249
QY 1219 GYQVKGAW 1226
Db 1250 GKLIKASY 1257

RESULT 14
US-09-995-749A-2
```

Db 1362 ANDPVSGLAVWVPVAGASNDQARTAPSTETKNSGNSAIRNAAFDNSVIFEFNSFYTP 1421
Qy 873 TKDSYTNKKIAQNVOLFKSWGVTSEMAPQYVSSSEDSGLDSIIQNGYAFEDRYDLAMS 932
Db 1422 TKESERANVRIAQNAADFFASLGFTSEMAPQYNSKDRDFTLDSIDNGYAFEDRYDLGMS 1481
Qy 933 KKKYSGSQDMINAVKALHSGI:QVIADWPDQIYNLPGEVVTATRVNDYGEYRKDSEI 992
Db 1482 EPNKYGTDEDRNAIQALHKGAGLQVMADWPDQIYNLPGEVATVTRVDDRGVWMDAI 1541
Qy 993 KNTLYAANTKSGNDYQAKYGGAFSLAELAKYPSIFNRTQISNGKKIDPSEKIKTANKAKY 1052
Db 1542 NNNLYVNTTGGG-EYQKYGGAFLDQKLQKLEIFTKQVSTGVADPSQKITESAKY 1600
Qy 1053 FNGTNILGRGVYLVKDNASDKYFELKGNQ--YLPKQMT--NKEASTGFV--NDGNGMT 1106
Db 1601 FNGTNILHRSGVYLVKADGG-QYVNL-GITTKQFLPIQLTGKKGQNEGFVKNDGN-YY 1657
Qy 1107 FYSTGYQAKNSFVQAKGNWYFFDNGHMYVGLQQLN---GE--VQYFSLNGVQLRES 1160
Db 1658 FYDLAGNMVKNFTIEDSVGNWYFFDQDGKVENKHFVDVDSYGEKGTYPFLKNGVSFRGG 1717
Qy 1161 FLENADGSKNYFCHLGNRYNSGYSPDNDKWKRYFDASGVMAVGLKTINGNTQYF--DQD 1218
Db 1718 LVQT-----DNGTYFFDNYGK-----WVRNQTIINAGAMIYTLDEN 1752
Qy 1219 GYQVKGAW 1226
Db 1753 GKLIKASY 1760

RESULT 15

US-09-604-957-5
; Sequence 5, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAQOI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 5
; LENGTH: 523
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-09-604-957-5

Query Match | 20.6%; Score 1534; DB 2; Length 523;
Best Local Similarity 55.6%; Pred. No. 1.7e-84;
Matches 303; Conservative 78; Mismatches 142; Indels 22; Gaps 3;

Qy 421 LLANDIDNSNPVQAEOLNHLNLYMNGYSIVANDPEANFDGVRVDAVNVNADLLQIASD 480
Db 1 LLANDVDNSNVVVEAEQLNWLNYLMNFGTITANDADANFDGIRVDAVNDVNDLQIAAD 60
Qy 481 YLKAHYGVDPKSEKNAIHLISLEAVSDNDPQYNKOTKGAQLPIDNKLRLSLYALTRPLE 540
Db 61 YKFLAYGVQDNDATANGHLISLEDSHNDPLVTDGSGNQLTMDQVYVHTQLIWSLTK--- 117
Qy 541 KDAKNKNEIRSGLEPVITNSLNNRAEGKNSERMANYIFIRAHDSVQTVIAKIIKAQIN 600
Db 118 -----SSDIRGTQMRFVDYVWVDRSNDSTENEAIPNYSFVRADSEVQTVIAQIVSDLYP 172
Qy 601 PXTDGLTFTLDELKQAFKLYNEDMQAKKYYTQSNIPYAYALMLNKSQITRLYYGDMYS 660
Db 173 DVENSLAPTEQLAAAFKYNEDKADKKYQYNMAYAMLLTNKOTVPRVYGDLYT 232

Qy 661 DDQYMATKSPYYDAIDTLLKARIKYAAGQDMKIITYVEGDGSHMDWDYTGVLTSVRYGT 720
Db 233 DDQYMATKSPYYDAIDTLLKARVOYVAGQSMV-----DSNDVLTSVRYGK 280
Qy 721 GANEATDQGESEATKTOGMVITSNPNPSLKLNDKQVIVNMGAAHKNQOEYRPLLLTTKQGL 780
Db 281 DAMTASDTCGTSETRTEGIVIVSNNAELQLEDGHTVTLHMGAHKNQAVRALLSTTADGL 340
Qy 781 TSVTSDAAAKSLYRKTNKNDKGLVFDASDIQGYLNPQVSGYLAWVPVVGASDNDQVRAAS 840
Db 341 AYTDDTENAPVAY--TDANGDLIFTNESYGVONPQVSGYLAWVPVGAQQDQDARTASD 398
Qy 841 NKANATGQYVYESSSALDSQLIYEGFSNFQDFVTKDSYTNKKIAQNVQLFKSMGVTSFEM 900
Db 399 TTTNTSDKVFHNSAALDSQVIYEGFSNFQAFATDSSEYTNVVTIAQNAQPKQWGTSPQL 458
Qy 901 APQYVSSSEDSGLDSIIQNGYAFEDRYDLAMSKNNKYGSQQDMINAVKALHKSQIQVIAD 960
Db 459 APQYRSSTDTSLDSIIQNGYAFEDRYDLGYPPTKYGTADQLRDAIKALHASGIIQAIAD 518
Qy 961 WVPDQ 965
Db 519 WVPDQ 523

Search completed: February 11, 2006, 19:45:46
Job time : 38.3088 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 20:39:23 ; Search time 127.994 Seconds
(without alignments)
4668.162 Million cell updates/sec

Title: US-10-797-821-36
Perfect score: 7462
Sequence: 1 METKRYKMKVKKHWTVA.....EGKQKGDVAYDERRLLVTR 1430

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA_Main:
1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pep.*
3: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pep.*
4: /cgn2_6/prodata/1/pubpaa/US10_PUBCOMB.pep.*
5: /cgn2_6/prodata/1/pubpaa/US10_PUBCOMB.pep.*
6: /cgn2_6/prodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7462	100.0	1430	3	US-09-740-274-6
2	7462	100.0	1430	4	US-10-383-930-36
3	7462	100.0	1430	5	US-10-797-821-36
4	3817.5	51.2	1475	3	US-09-740-274-2
5	3817.5	51.2	1475	4	US-10-383-930-34
6	3817.5	51.2	1475	5	US-10-797-821-34
7	3615	48.4	1554	4	US-10-383-930-38
8	3615	48.4	1554	5	US-10-797-821-38
9	3591.5	48.1	1375	3	US-09-740-274-4
10	3591.5	48.1	1375	4	US-10-383-930-35
11	3591.5	48.1	1375	5	US-10-797-821-35
12	3575	47.9	1590	4	US-10-383-930-37
13	3575	47.9	1590	5	US-10-797-821-37
14	3502.5	46.9	1518	4	US-10-383-930-40
15	3502.5	46.9	1518	5	US-10-797-821-40
16	3364.5	45.1	1497	5	US-10-484-218-18
17	3174	42.5	1365	4	US-10-383-930-39
18	3174	42.5	1365	5	US-10-797-821-39
19	3087	41.1	1595	3	US-10-484-218-20
20	2780	37.3	545	3	US-09-995-749A-10
21	2699.5	36.2	2057	4	US-10-417-280A-2
22	2460.5	33.0	1777	5	US-10-484-218-12
23	2446	32.8	1781	3	US-09-995-749A-2
24	2380.5	31.9	1771	5	US-10-484-218-14
25	2349	31.5	1006	5	US-10-484-218-22
26	1532.5	20.5	522	3	US-09-995-749A-11
27	1400	18.8	535	3	US-09-995-749A-13

28	1306.5	17.5	584	3	US-09-995-749A-12	Sequence 12, Appl
29	1257.5	16.9	787	5	US-10-484-218-16	Sequence 16, Appl
30	1246	16.7	525	5	US-10-484-218-23	Sequence 23, Appl
31	668.5	9.0	224	5	US-10-484-218-4	Sequence 4, Appl
32	664	8.9	223	5	US-10-484-218-10	Sequence 10, Appl
33	647	8.7	223	5	US-10-484-218-6	Sequence 6, Appl
34	515	6.9	221	5	US-10-484-218-8	Sequence 8, Appl
35	509	6.8	221	5	US-10-484-218-2	Sequence 2, Appl
36	453	6.1	2710	4	US-10-011-366-6	Sequence 6, Appl
37	453	6.1	2710	4	US-10-354-774-6	Sequence 6, Appl
38	453	6.1	2710	4	US-10-271-012-6	Sequence 6, Appl
39	453	6.1	2710	4	US-10-729-122-6	Sequence 6, Appl
40	453	6.1	2710	4	US-10-729-039-6	Sequence 6, Appl
41	453	6.1	2710	5	US-10-729-527-6	Sequence 6, Appl
42	453	6.1	2710	5	US-10-727-898-6	Sequence 6, Appl
43	453	6.1	2710	5	US-10-728-696-6	Sequence 6, Appl
44	453	6.1	2710	6	US-11-001-241-6	Sequence 6, Appl
45	387	5.2	2366	4	US-10-011-366-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-09-740-274-6
; Sequence 6, Application US/09740274
; Patent No. US30020031826A1
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-6

Query Match	100.0%;	Score 7462;	DB 3;	Length 1430;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1430;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	METKRYKMKVKKHWTVA	ASGLITLGTTLTGSSVSAETEQTS	SDKVVTKSEDDKAA 60
Db	1	METKRYKMKVKKHWTVA	ASGLITLGTTLTGSSVSAETEQTS	SDKVVTKSEDDKAA 60
QY	61	SSSQTDAPKTQKQAEQTQ	QAQANVADTSTSI	TKETPSQNTTQANSDDKVTNTKSE 120
Db	61	SSSQTDAPKTQKQAEQTQ	QAQANVADTSTSI	TKETPSQNTTQANSDDKVTNTKSE 120
QY	121	EAQTSEERTKQSEEAQT	TASSQALTOAKAELTKQRTA	QAKENKPNVDLAAIPNVKQIDGK 180
Db	121	EAQTSEERTKQSEEAQT	TASSQALTOAKAELTKQRTA	QAKENKPNVDLAAIPNVKQIDGK 180
QY	181	YYVIGSDGQPKKNFAL	TNNKVLVYFDKNTGALT	DTTSQYQFKGLTKLNNDYTPHNOIVNF 240

Db 181 YYYIGSDGQPKKNFALTNNKVLVPDKNTGALTDTTSQYQKQGLTKLNNNDYTPHNIQVNF 240
Qy 241 ENTSLETIDNYTADSWYRPPKDLKNGKTWTWASSSDLRPLLMSWMPDKQTOIAYLNNYN 300
Db 241 ENTSLETIDNYTADSWYRPPKDLKNGKTWTWASSSDLRPLLMSWMPDKQTOIAYLNNYN 300
Qy 301 QQGLGTGENYTADSSQESLNLAQTQVQKIETKISQTOQTQWLRLDINSFVKTPQNNNSQ 360
Db 301 QQGLGTGENYTADSSQESLNLAQTQVQKIETKISQTOQTQWLRLDINSFVKTPQNNNSQ 360
Qy 361 TESDTSAGEKDLQGLGALLYSNSDKTAYANSYRLLNRTPTSTQTKPKYFEDNSSGGYDF 420
Db 361 TESDTSAGEKDLQGLGALLYSNSDKTAYANSYRLLNRTPTSTQTKPKYFEDNSSGGYDF 420
Qy 421 LLANDIDNSNPVQAEQNLWHLNMYGSIIVANDPEANFDGVRVDAVDNVDNADLLQIASD 480
Db 421 LLANDIDNSNPVQAEQNLWHLNMYGSIIVANDPEANFDGVRVDAVDNVDNADLLQIASD 480
Qy 481 YLKAHGVVDKSEKNAIHLNLSILEAWSDNDPQYNKDTKGALPDKLRLSLIYALTRPLE 540
Db 481 YLKAHGVVDKSEKNAIHLNLSILEAWSDNDPQYNKDTKGALPDKLRLSLIYALTRPLE 540
Qy 541 KDAKNEIRSGLEPVITNSLNNRSGKNSERMANYIFIRAHDSVQTVIAKIIKAQIN 600
Db 541 KDAKNEIRSGLEPVITNSLNNRSGKNSERMANYIFIRAHDSVQTVIAKIIKAQIN 600
Qy 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYYTQSNIPTAYALMLSNKDSITRLYGDWYS 660
Db 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYYTQSNIPTAYALMLSNKDSITRLYGDWYS 660
Qy 661 DPGQYMATKSPYYDAIDTLKARIKYAAGQDMKITYVEGDKSHMDWDYTGVLTSVRYCT 720
Db 661 DPGQYMATKSPYYDAIDTLKARIKYAAGQDMKITYVEGDKSHMDWDYTGVLTSVRYCT 720
Qy 721 GANEATDQGESEATKTQGMVITSNPDLKLNQNDKVIIVNMGAAHKNQYRPLLLTKQGL 780
Db 721 GANEATDQGESEATKTQGMVITSNPDLKLNQNDKVIIVNMGAAHKNQYRPLLLTKQGL 780
Qy 781 TSYTSDAAKSILYRKTNDKGEIVFDASDIQGYLNPQVSGYLAVVVPVGASDNQDVRVAAS 840
Db 781 TSYTSDAAKSILYRKTNDKGEIVFDASDIQGYLNPQVSGYLAVVVPVGASDNQDVRVAAS 840
Qy 841 NKANATGOVYESSALDSOLIVYEGFSNFODFTVKDSYTNKKTAAQNVQLPKSNGWTSFEM 900
Db 841 NKANATGOVYESSALDSOLIVYEGFSNFODFTVKDSYTNKKTAAQNVQLPKSNGWTSFEM 900
Qy 901 APQYVSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSQDMINAVKALHKSIGIQVIAD 960
Db 901 APQYVSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSQDMINAVKALHKSIGIQVIAD 960
Qy 961 WVPDQIYNLPKGEVVTATRVNDYGEYRKDSEIKNTLYAANTKSNKGDKYQAKYGGAPLSEL 1020
Db 961 WVPDQIYNLPKGEVVTATRVNDYGEYRKDSEIKNTLYAANTKSNKGDKYQAKYGGAPLSEL 1020
Qy 1021 AAKYPSIFNRKTOISNGKKIDPSEKITAWKAKYFNGNIIILGRGVYVLDKNASDKYFELKG 1080
Db 1021 AAKYPSIFNRKTOISNGKKIDPSEKITAWKAKYFNGNIIILGRGVYVLDKNASDKYFELKG 1080
Qy 1081 NOTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYFDNNGHMVYGL 1140
Db 1081 NOTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYFDNNGHMVYGL 1140
Qy 1141 QQLNGEVQYFLSNGVOLRESFLENADGSKNYFGHLNRYSNGYSPDNDSKWRYPDASGV 1200
Db 1141 QQLNGEVQYFLSNGVOLRESFLENADGSKNYFGHLNRYSNGYSPDNDSKWRYPDASGV 1200
Qy 1201 MAVGLKTINGNTQYFPDQDGQVQKAWITGSDGKKRYFDDGSGNMAVNRFANDKNGDWYYL 1260
Db 1201 MAVGLKTINGNTQYFPDQDGQVQKAWITGSDGKKRYFDDGSGNMAVNRFANDKNGDWYYL 1260
Qy 1261 NSDGIALVGQTINGNTKYTFYFGDQKQIKGKIITDNGKLYFLANSSELARNIPATDSQNN 1320
Db 1261 NSDGIALVGQTINGNTKYTFYFGDQKQIKGKIITDNGKLYFLANSSELARNIPATDSQNN 1320

Qy 1321 WYFEGSDGVAVTGSQTIAGKKLYPASDQGVKSGSVTYNGKVHYVHADSGELQVNRFEAD 1380
Db 1321 WYFEGSDGVAVTGSQTIAGKKLYPASDQGVKSGSVTYNGKVHYVHADSGELQVNRFEAD 1380
Qy 1381 KDGNNYYLDSNGEALTGSRINDQRVFFTRREGKQVKGVDVAYDERLLVYR 1430
Db 1381 KDGNNYYLDSNGEALTGSRINDQRVFFTRREGKQVKGVDVAYDERLLVYR 1430

RESULT 2

US-10-383-930-36
; Sequence 36, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 36
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-36

Query Match 100.0%; Score 7462; DB 4; Length 1430;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 METKERYKMHKVKHMTVAVASGLITLGTTLTGSSVSAETEQOTSDKVVTOKSDDKAA 60
Db 1 METKERYKMHKVKHMTVAVASGLITLGTTLTGSSVSAETEQOTSDKVVTOKSDDKAA 60
Qy 61 SSSQTDAPKTKQAOTEQTQAOSQANVADTSTSIKTETPSQNTTQAASDDKTVNTKSE 120
Db 61 SSSQTDAPKTKQAOTEQTQAOSQANVADTSTSIKTETPSQNTTQAASDDKTVNTKSE 120
Qy 121 EAQTSSEERTKQSEEAQTASSQALTOAKAELTKQRTAAQENKNPVDLAAIPNVKQIDGK 180
Db 121 EAQTSSEERTKQSEEAQTASSQALTOAKAELTKQRTAAQENKNPVDLAAIPNVKQIDGK 180
Qy 181 YYYIGSDGQPKKNFALTNNKVLVPDKNTGALTDTTSQYQKQGLTKLNNNDYTPHNIQVNF 240
Db 181 YYYIGSDGQPKKNFALTNNKVLVPDKNTGALTDTTSQYQKQGLTKLNNNDYTPHNIQVNF 240
Qy 241 ENTSLETIDNYTADSWYRPPKDLKNGKTWTWASSSDLRPLLMSWMPDKQTOIAYLNNYN 300
Db 241 ENTSLETIDNYTADSWYRPPKDLKNGKTWTWASSSDLRPLLMSWMPDKQTOIAYLNNYN 300
Qy 301 QQGLGTGENYTADSSQESLNLAQTQVQKIETKISQTOQTQWLRLDINSFVKTPQNNNSQ 360
Db 301 QQGLGTGENYTADSSQESLNLAQTQVQKIETKISQTOQTQWLRLDINSFVKTPQNNNSQ 360
Qy 361 TESDTSAGEKDLQGLGALLYSNSDKTAYANSYRLLNRTPTSTQTKPKYFEDNSSGGYDF 420
Db 361 TESDTSAGEKDLQGLGALLYSNSDKTAYANSYRLLNRTPTSTQTKPKYFEDNSSGGYDF 420
Qy 421 LLANDIDNSNPVQAEQNLWHLNMYGSIIVANDPEANFDGVRVDAVDNVDNADLLQIASD 480
Db 421 LLANDIDNSNPVQAEQNLWHLNMYGSIIVANDPEANFDGVRVDAVDNVDNADLLQIASD 480
Qy 481 YLKAHGVVDKSEKNAIHLNLSILEAWSDNDPQYNKDTKGALPDKLRLSLIYALTRPLE 540
Db 481 YLKAHGVVDKSEKNAIHLNLSILEAWSDNDPQYNKDTKGALPDKLRLSLIYALTRPLE 540

QY 541 KDAASNKNEIRSGLEPVITNSLNRSABGKNSRMANYIFIRAHDSVQTVIAKIIKAQIN 600
DB 541 KDAASNKNEIRSGLEPVITNSLNRSABGKNSRMANYIFIRAHDSVQTVIAKIIKAQIN 600
QY 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYTOSNIPTAYALMLSNKDSITRLYYGDMYS 660
DB 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYTOSNIPTAYALMLSNKDSITRLYYGDMYS 660
QY 661 DDQGYMATKSPYYDAIDTLKARIKYAAGQDMKITYVEGDKSHMDWDYTGVLTSVRYGT 720
DB 661 DDQGYMATKSPYYDAIDTLKARIKYAAGQDMKITYVEGDKSHMDWDYTGVLTSVRYGT 720
QY 721 GANEATDQSEATKTQGMVITNSNPSLKLNDKQIVNMGAAHKNQYRPLLLTKDGL 780
DB 721 GANEATDQSEATKTQGMVITNSNPSLKLNDKQIVNMGAAHKNQYRPLLLTKDGL 780
QY 781 TSVTSDAAKSLYRKNTDKGELVFDASDIQGYLNPQVSGYLA VMPVPGASDNODVRAAS 840
DB 781 TSVTSDAAKSLYRKNTDKGELVFDASDIQGYLNPQVSGYLA VMPVPGASDNODVRAAS 840
QY 841 NKANATQGVYESSALDSQLIYEGFSNFQDFTVKDSYTNKKIAQNVQLFKSWGVTSEFM 900
DB 841 NKANATQGVYESSALDSQLIYEGFSNFQDFTVKDSYTNKKIAQNVQLFKSWGVTSEFM 900
QY 901 APOYVSEDSGLDLSIIQNGYAFEDRYDLAMSKNKYGSQDMINAVKALHKSIGIOVIAD 960
DB 901 APOYVSEDSGLDLSIIQNGYAFEDRYDLAMSKNKYGSQDMINAVKALHKSIGIOVIAD 960
QY 961 WPDQIYNLPGKEVVTATRVNDYGEYRKDSEIKNTLYAANTKSNKGDKYQAKYGGAFSEL 1020
DB 961 WPDQIYNLPGKEVVTATRVNDYGEYRKDSEIKNTLYAANTKSNKGDKYQAKYGGAFSEL 1020
QY 1021 AAKYPSIFNRTQISNGKKIDPSEKITA WAKAKYFNGTILGRGVYVLDKNASDKYFELKG 1080
DB 1021 AAKYPSIFNRTQISNGKKIDPSEKITA WAKAKYFNGTILGRGVYVLDKNASDKYFELKG 1080
QY 1081 NOTYLPKQMTNKEASTGFVNDGNGMTFYSTGSOAKNSFVQDAKGNWYFDNNGHVYGL 1140
DB 1081 NOTYLPKQMTNKEASTGFVNDGNGMTFYSTGSOAKNSFVQDAKGNWYFDNNGHVYGL 1140
QY 1141 QOLNGEVOYFLSNGVQLRESFLENADGSKNYFCHLGNYSNGYYSFDNDSKRYFPDASGV 1200
DB 1141 QOLNGEVOYFLSNGVQLRESFLENADGSKNYFCHLGNYSNGYYSFDNDSKRYFPDASGV 1200
QY 1201 MAVGLKTINGNTQYFQDQGVQVKGAWITGSDGKRYFDDGSGNMAVNRFANDKNGDWYVL 1260
DB 1201 MAVGLKTINGNTQYFQDQGVQVKGAWITGSDGKRYFDDGSGNMAVNRFANDKNGDWYVL 1260
QY 1261 NSDGIALVGVQTINGKTYTFYFGDGKQIKGKIITDNGKLYFLANSSELARNIFATDSQNN 1320
DB 1261 NSDGIALVGVQTINGKTYTFYFGDGKQIKGKIITDNGKLYFLANSSELARNIFATDSQNN 1320
QY 1321 WYFEGSDVAVTGSQTIAGKLYFASDGQVKGFSFVYNGKHYHADSSELQVNRFEAD 1380
DB 1321 WYFEGSDVAVTGSQTIAGKLYFASDGQVKGFSFVYNGKHYHADSSELQVNRFEAD 1380
QY 1381 KQGNWYILDSNGEALTGSRINDQRFVFFREGQVKGVDVAYDERLLVTR 1430
DB 1381 KQGNWYILDSNGEALTGSRINDQRFVFFREGQVKGVDVAYDERLLVTR 1430

RESULT 3
US-10-797-821-36
; Sequence 36, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930

; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 36
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-36

Query Match 100.0%; Score 7462; DB 5; Length 1430;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 METKRYKMHKVKHWHVTVAVASGLITLGTTLTGSSVSAETEQTSDKVVTKSEDDKAA 60
DB 1 METKRYKMHKVKHWHVTVAVASGLITLGTTLTGSSVSAETEQTSDKVVTKSEDDKAA 60
QY 61 SSSOTDAPKTKQAQTEQTAQSOANVADTSTITKETPSQNTTQANSDDKTVTWTKE 120
DB 61 SSSOTDAPKTKQAQTEQTAQSOANVADTSTITKETPSQNTTQANSDDKTVTWTKE 120
QY 121 EAQTSERTKQSEEAQTASSQALTOAKAELTKQRTAAQENKPNVDLAAIPNVKQIDGK 180
DB 121 EAQTSERTKQSEEAQTASSQALTOAKAELTKQRTAAQENKPNVDLAAIPNVKQIDGK 180
QY 181 YYYIGSDGPKKNFALTVNNKVLVFDKNTGALTDTSTQYQFKGLTKLNDYTPHNIQVNF 240
DB 181 YYYIGSDGPKKNFALTVNNKVLVFDKNTGALTDTSTQYQFKGLTKLNDYTPHNIQVNF 240
QY 241 ENTSLETIDNYTADSWYRPKDIKNGKWTWTSSESDELPLLMWPKQTOIAYLNYN 300
DB 241 ENTSLETIDNYTADSWYRPKDIKNGKWTWTSSESDELPLLMWPKQTOIAYLNYN 300
QY 301 QOGLGTGENYATDSQESLNLAAQTQVVKIETKISOTQOTQWLRIINSFVKTPQNNNSQ 360
DB 301 QOGLGTGENYATDSQESLNLAAQTQVVKIETKISOTQOTQWLRIINSFVKTPQNNNSQ 360
QY 361 TESDTSAGEKHLQGGALLYNSDKTAYANSYRLLNRTPTSTQTKPKYFEDNSSGGYDF 420
DB 361 TESDTSAGEKHLQGGALLYNSDKTAYANSYRLLNRTPTSTQTKPKYFEDNSSGGYDF 420
QY 421 LLANDIDNSNPVQAEQNLWHLVNYGSIIVANDPEANFDGVRVADVNDVNDLLOIASD 480
DB 421 LLANDIDNSNPVQAEQNLWHLVNYGSIIVANDPEANFDGVRVADVNDVNDLLOIASD 480
QY 481 YLKAHYGVDKSEKNAINHLSILEANSNDPQNKDTKGALPDKLRLSLLYALTRPLE 540
DB 481 YLKAHYGVDKSEKNAINHLSILEANSNDPQNKDTKGALPDKLRLSLLYALTRPLE 540
QY 541 KDAASNKNEIRSGLEPVITNSLNRSABGKNSRMANYIFIRAHDSVQTVIAKIIKAQIN 600
DB 541 KDAASNKNEIRSGLEPVITNSLNRSABGKNSRMANYIFIRAHDSVQTVIAKIIKAQIN 600
QY 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYTOSNIPTAYALMLSNKDSITRLYYGDMYS 660
DB 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYTOSNIPTAYALMLSNKDSITRLYYGDMYS 660
QY 661 DDQGYMATKSPYYDAIDTLKARIKYAAGQDMKITYVEGDKSHMDWDYTGVLTSVRYGT 720
DB 661 DDQGYMATKSPYYDAIDTLKARIKYAAGQDMKITYVEGDKSHMDWDYTGVLTSVRYGT 720
QY 721 GANEATDQSEATKTQGMVITNSNPSLKLNDKQIVNMGAAHKNQYRPLLLTKDGL 780

Db 721 GANEATDQGEATKTOGMAVITSNPDLKLNQNDKVINVMGAHKNQYRPLLLTTKQGL 780
QY 781 TSYTSDAAKSLYRKNTDNGELVFDASDIQGYLNPQVSGYLAVVVPVGASDNQDVRVAAS 840
Db 781 TSYTSDAAKSLYRKNTDNGELVFDASDIQGYLNPQVSGYLAVVVPVGASDNQDVRVAAS 840
QY 841 NKANATGOVYESSALDSOLIYEGFSNFODFVTKSDYTNKKTIAQNVQLFKSGWTSPEM 900
Db 841 NKANATGOVYESSALDSOLIYEGFSNFODFVTKSDYTNKKTIAQNVQLFKSGWTSPEM 900
QY 901 APOYVSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSQQDMINAVKALHKSGLQVIAD 960
Db 901 APOYVSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSQQDMINAVKALHKSGLQVIAD 960
QY 961 WYPDQIYNLPKGEVVTATRVNDYGEYKDSIKNTLYAANTKSGNKDYQAKYGGAPLSEL 1020
Db 961 WYPDQIYNLPKGEVVTATRVNDYGEYKDSIKNTLYAANTKSGNKDYQAKYGGAPLSEL 1020
QY 1021 AAKYPSIFNRTQISNGKKIDPSEKITAWKAKYFNGTNIILGRGVGYVLKDNASDKYFELKG 1080
Db 1021 AAKYPSIFNRTQISNGKKIDPSEKITAWKAKYFNGTNIILGRGVGYVLKDNASDKYFELKG 1080
QY 1081 NOTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYFDNNGHMVYGL 1140
Db 1081 NOTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYFDNNGHMVYGL 1140
QY 1141 QQLNGEVQVFLSNGVOLRESFLENADGSKNYFGLHGNRYSGNYSPDNDSKWRYFDASGV 1200
Db 1141 QQLNGEVQVFLSNGVOLRESFLENADGSKNYFGLHGNRYSGNYSPDNDSKWRYFDASGV 1200
QY 1201 MAVGLKTINGTQYFDDGQYQVKGAMITGSDGKKRYFDDGSGNMAVNRANDKNGDWYVL 1260
Db 1201 MAVGLKTINGTQYFDDGQYQVKGAMITGSDGKKRYFDDGSGNMAVNRANDKNGDWYVL 1260
QY 1261 NSDGIALVGQTINGKTYTFYFGQDKQIKGKIITDNGKLYFLANSSELARNIFATDSQNN 1320
Db 1261 NSDGIALVGQTINGKTYTFYFGQDKQIKGKIITDNGKLYFLANSSELARNIFATDSQNN 1320
QY 1321 WYFYGSDGVAVTGSQTIAGKLYFASDGQVKGSGFVYTKGVHYHADSGELQVNRFEAD 1380
Db 1321 WYFYGSDGVAVTGSQTIAGKLYFASDGQVKGSGFVYTKGVHYHADSGELQVNRFEAD 1380
QY 1381 KQGNWYFLDSNGEALTGSORINDORVFFTRGKQVKGDVAYDERRLLVYR 1430
Db 1381 KQGNWYFLDSNGEALTGSORINDORVFFTRGKQVKGDVAYDERRLLVYR 1430

RESULT 4

US-09-740-274-2
; Sequence 2, Application US/09740274
; Patent No. US20020031826A1
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740, 274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-740-274-2

Query Match 51.2%; Score 3817.5; DB 3; Length 1475;
Best Local Similarity 52.9%; Pred. No. 9.7e-190; Mismatches 395; Indels 53; Gaps 22;
Matches 762; Conservative 230;
QY 1 METKERYKMHKVKKHWTVAVASGLITLTGTTTGLSSVSAETBQOTSQDKVVTQKSEDDKAA 60
Db 1 MDKVRVYKLRKVKKRWTVSVASA VMTL-TTLLSGGLVKADSNESKS-----QISNDSNTS 54
QY 61 SSSQTDAPKTKQAOTEQTOAQSOAN-VADTSTSTITKETPSQNIITTOANSDDKTVTNTKS 119
Db 55 VVTANEESNVITEATSKQEAASSQTNHTVTTSSTSTSVVNPKEVSNPYTVGTETASNGEK 114
QY 120 BEAOTSSEBRTKQSEEAQTTFASSQALTQAKAELTKQRTAAQENKPNVDLAAIPNVKOIDG 179
Db 115 LQNTQITTV-DKTSAAANNISKQT-TEADTVDDSNAA-----NLQILEKLPNVKEIDG 167
QY 180 KYIYIGSDGQPKKNFALTNNKVLKYFDKNTGALTDTTS-QYQPKQGLTKLNNDYTPHNQIV 238
Db 168 KYIYVYDNNKGVRTNFTLIADGKILHFDE-TGAYTDTSDITVNKDIVTTRSNLYKKNQYV 226
QY 239 NFENTSLETIDNYVTADSWYRPKDILKNGKWTASSEDLRLPLMSWHPDKQTOAYLNY 298
Db 227 DRSAQSFHFVDHYLTAESWYRPKYILKDGKTWTQSTEKDFRPLMTWMPDQSTQRYNY 286
QY 299 MNQOGLGTGENTYADSSQESLNLAQTVQVKLETKISOTQOTQWLRLDIINSFVKTOPNN 358
Db 287 MNAQ-LGINKYVDDTSNQLNLAAATQAKTEAKITLLKNTDMLRQTISAFVKTSQAWN 345
QY 359 SOTESDTSAGEKHLQGGALLYNSDK-TAYANSYRLNLRPTTSQTK--PKYFEDNSS 415
Db 346 SDSEKPPD----DHLQNGAVLYDNEGKLTTPYANSYRILNRPNTQTKKDPRTADNTI 401
QY 416 GGYDFLLANDINSNPVVOAEQLNWLHLYMNYGSIIVANDPEANFGVRVDVNDVNADLL 475
Db 402 GGYEFLANDVDSNPVVOAEQLNWLHFLMNFNGIYVANDPDANFDSIRVDAVDNVDADLL 461
QY 476 QIASDYLKAHYGVDSKSEKNAINHLSILEAWSNDQYKNTKGAOLPDKNLRLSLLYAL 535
Db 462 QIAGDYLKAAKGIHNDKANDHLSILEAWSNDTPYLDHDDGDNMINMDNKLRLSLFL 521
QY 536 TRPLEKASNKNEIRSGLEPVTISLNNRSASGKNSERMANYIFIRAHDSVQTVIAKII 595
Db 522 AKPLNQ-----RSGMNPITNSLVNRTDDNAETAAPVPSYFIRAHDSVQDLIADI 573
QY 596 KAOINPKTDGLTFTDELKQAFKINEDMRQAKKYTQSNITPVALMLSNKDSITRLYY 655
Db 574 KAEINPNVVGYSFTMEIEKKAPEIYNKOLLATEKKYTHYNTALSYALLLTKNSSVPRVY 633
QY 656 GDMYSDDGOYMATKSPYDAITLKLKARIKYAAGQDMKIITYVEGDKSHMDWDTYCVLTS 715
Db 634 GDMFTDDGQYMAKHTINYEAIETLKLARIKYVGGQMRNQVGNSE-----IITIS 684
QY 716 VRYGTGANEATDQGEATKTOGMAVITSNPDLKLNQNDKVINVMGAHKNQYRPLLLT 775
Db 685 VRYGKALKATDTGDRTRTSGAVIEGNNPSLRKASDRVVVMGAHKNQYRPLLLT 744
QY 776 TKDGLTSYTSDAKASLRYKTNKDELVPDASDIQGYLNPQVSGYLAVVVPVGASDNQDV 835
Db 745 TDNGIKAVHSDQEAAGLVRYTNDRGELIFTAADIKGYANPQVSGYLGWVPVGA---LI 801
QY 836 RVAASNKANATGV--YESSALDSOLIYEGFSNFODFVTKSDYTNKKTIAQNVQLFKSW 893
Db 802 KMFALRLARPHQOQMASVHQNAALDSKVMFEFSNQAFAKKEEYTNVVIKNDVFAEW 861
QY 894 GYTSFEMAPQYVSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSQQDMINAVKALHKS 953
Db 862 GYTDfEMAPQYVSSDGSFLDSVIQNGYAFTRDYDLGISKPNKYGTADDLVKAIKALHKS 921


```

Db      1219  NEEGKWLVDGEDGAAVTGSQTINGQHLYFRANGVQVKGEFVTDHGRISYYDGNSGDQIR 1278
QY      1311  NIEATDSQNNWYFGSDGVAVTQSOTIAGKKLYFASDKQVKGSFVTYN-GKVHYIHADS 1369
Db      1279  NRFVRNAQGMWFYDNNGYAVTGARTINGQLLYFRANGVQVKGEFVTDYGRISYYDGN 1338
QY      1370  GEIQVNRFEADKQGNWYLYDSNGEALTGSQRINDQRFVFTREGQVKGVAYDERLLVY 1429
Db      1339  GDQIRNRFVRNAQGMWFYDNNGYAVTGARTINGQHLYFRANGVQVKGEFVTDHGRISY 1398

RESULT 6
US-10-797-821-34
; Sequence 34, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Taubman, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogen
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 34
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-34

```

346	Db	SDEKPD-----DHLQNGAVLYONEGKLTPTYANSNVRILNRTPTNQTKKDPRTADNTI	401
416	Qy	GGYDFLLANDINSNPVQAEQLNWLHLYLMNYGSIVANDPEANFDGVRVDAVNVNADLL	475
402	Db	GGYEFLLVANDVNSNPVQAEQLNWLHFLMNFGNIVANDPDANFDSIRVDAVNVNADALL	461
476	Qy	QIASDYLLKAHYGVDDKSEKNAINHLSILEAWSNDNDPQYNKDTKGAQLPIDNKLRLSLLYAL	535
462	Db	QIAGDYLLKAAGIKHNDKAANDHLSILEAWSNDNTPYLDHDDGDNMIMNDKJLSLFSL	521
536	Qy	TRPLEKASDNKNEIRSGLEPVTINSNLNRSABGKNSERMANYLIFIRAHDSVQTVIAKII	595
522	Db	AKPLNQ-----RSGMNLITNSLVNRTRDDNAETAAPSYGSFIRAHDSVQDLADII	573
596	Qy	KAQINPKTDGLTFLDBLKQAFKIYNEDMRQAKKYTQSNIPATAYALMLSNKDSITRLYY	655
574	Db	KAEINPNVGVSTMEBIIKKAPELYNKDALLATEKKYTHYNTALSYALLLTNKSSVPRVY	633
656	Qy	GDWYSDGQYMAKSPYYDAIDLTKKARIKYAAGQDMKITYYEGDKSHMDWDYTGVLTS	715
634	Db	GDMPDGDQYMAHKTINYEAETLLKARIKYVSGQAMRNQOVGNSE-----IITS	684
716	Qy	VRYCTGANENATDOGSEATKTQGMAVITSNPSSLKLNQNDKVIYNMGAHKNQOERPLLLT	775
685	Db	VRYGKGALKATDGTDRTRTSGVAVIEGNNPSURLKASDRVVNMGAHKNQAYRPLLLT	744
776	Qy	TKDGLTSYTDAAKSLRYKTNDKGELVFOASDIQGYLNPQVSGYLAVMVVPVQASDNQDV	835
745	Db	TDNGIKAYHSDQEAAGLVRVYTNDRGELIFTAADIKGYANPQVSGYLGWVPVGNA---LI	801
836	Qy	RVAASNKANATQGV--YESSALSLSLIYEGFSNFPQDVFVKDSYTNKKIAQNVQLPKSW	893
802	Db	KMPALRLARPHQQWASVHQNAALDSRVMPGSGSNFQAFATKCEBYTNVIAKNVDKFAEW	861
894	Qy	GVTSFEMAPQVSESDGSFLDSIIIONGYAFEDRYDLAMSKNKYKGSQODMINAVKALHS	953
862	Db	GVTFEMAPQVASTDGSFLDSVIQNGYAFTRDYDLGISKPNKYGTADDDILKAIFALHSK	921
954	Qy	GIQVIADWVPDQIYNLPGKEVWTRVNDYGEYKDSSEIKNTLYAAMTKNGKDYQAKYG	1013
922	Db	GIRVMAWVPDQMTAFPEKEVWTRVDTKGTVPVAGSQIKNTLYVVDGKSGKDOQAKYG	981
1014	Qy	GAPLSLAAKYPISFINRTOISNGKKIDPSEKITAWKAKYFNGTNILGRGVGYVLKDNASD	1073
982	Db	GAPLEELQAKYPPELFARKQISTGVPMDDPSVKIKQWSAKYFNGTNILGRGAGYVLKDOQTN	1041
1074	Qy	KYPELKQNG--TYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAGNWTYFD	1131
1042	Db	TYFNISDNKEINFUPLKTYLLMQVSGFSYDGKGYVYVYSTGYQAKNTFISEG-DKWYTFD	1100
1132	Qy	NNGHMYGLQOLNGEVOYFELNSGVQLRESFLENADGSKNYFHLGNYSNGYYSFDDNSK	1191
1101	Db	NNGFMVTVGASINGVNTYFLSNGLQLKDALILKNEHDGTYAYGNDGRYENGYQF-MSGV	1159
1192	Qy	WRYPDASGVMAVGLTKTINGNTQYFDQGYQVKGAWITGSDGKKRYFDGSGNMAVNRPAN	1251
1160	Db	WRHEN-NGEMSUGLTVIDGQVQYFDEMGYQAKGKFVTADGKIRYPDKSGNMYNRNFIE	1218
1252	Qy	DKNGDYWLNSDGIALVGVOITNGKTYVFGDGQKQIKGKIITD-NGKLKTYFLANSGEJAR	1310
1219	Db	NEEGKWLTYLGEDGAAVTGSQTINGHLYFRANGVQVKGFEFTDHHGRISYIDYDNGSGDQIR	1278
1311	Qy	NI FATDSQNNWYFGSDGVAVTGSQTAGKLYFASDGKVKGSFVTYN-GKVHYVHAD5	1369
1279	Db	NRFVRNAQGMQYFPDNGYAVTGARTINGQLLYFRANGVQVKGFEFTDRYGRISYIDGNS	1338
1370	Qy	GELQVNRFEADKQGNWYTLDSNGBALGTSORINDQRVFFREGQVQKGVDAYDBRRLLVY	1429
1339	Db	GDQJNRFVRNAQGMQYFPDNGYAVTGARTINGHLYFRANGVQVKGFEFTDRHGRISY	1398

RESULT 7
US-10-383-930-38

```
; Sequence 38, Application US/10383930 /
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Taubman, Daniel J
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 38
; LENGTH: 1554
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
; US-10-383-930-38

Query Match      48.4%; Score 3615; DB 4; Length 1554;
Best Local Similarity 49.9%; Pred. No. 3.5e-179;
Matches 745; Conservative 217; Mismatches 418; Indels 112; Gaps 32;

QY      1  METKRYKMHKVKHWVTVAVAS-GLITL---GTTLLGSSVSABEQOTSDKYVTQKSED 56
      1  MEKKLHYKHLKHKVHWVTIAVASIGLVSLVGAGTVAEDKVANDTTAQTATGVGDTQ--- 57
QY      57  DKAAESSQTDAPKT-----KQATEQTOAQS-QANVADTSTSTIKETPTQNITTOANSDD 111
      58  DQATTNDANTNTDTPADQSANNTNQDQSGDSNNQDQAKQDT-----ANTDR 106
QY      112  KTWNTKSEAOETSEBRTKQSEBAQTASSQALTOAKAELTKOROTAAQENKPNVDLAAI 171
      107  NQADNSQTNQATDQATSPATDGTQVQRDAANVATA-ADQEGQTAPSEQEKSAAL-SL 164
QY      172  PNVKQIDGKYIYIGSDGPKQNFALTVNNKVLVYFDKNTGALTDTQYQFKQGLTKLNNDY 231
      165  DNVLKIDGKYIYVQADSGYKKNFAITVNGQMLYFDSDTGALSTSTYSFSQGTNLVDVF 224
QY      232  THNQIVNFENTSLFTIDNYVTADSWYRPKDIKNGKTWTASSESLRPLILMSWPDKOT 291
      225  SSKNAYDSTAKSFELVNGYLTANSWYRPAGILRNQGTWEASNENDLRPVLMSWPDKOT 284
QY      292  QIAYLNMNQOGLGTGENTYADSSQESNLAAQTVQVKLETKISQTOQTOWLRDIINSFV 351
      285  QVAYVNMNKYLSANETEVTNETSQVDLNKEAQSIQTKIEQKITSDNSTOWLRTAMEAFV 344
QY      352  KTQPNWNSQTESDTSAGEKDHLOGGALLYSNSDKTAYANSYRLNRTPTSTQTKPKYPE 411
      345  AAQPKWNMSTE---NPNKGDLHLOGGALLYTNSDLTPWANSYRLNRTPTQDGTCKEYF 401
QY      412  DNSSGGYDFLLANDTNSNPVQAEQLNHLVLMYLVGSIIVANDPEANFDGVRVADVNVN 471
      402  EGGEGRYEPLNSDNDNSNPVQAEQLNHLVLMWGDIVMGDKDANFDGVRVADVNVN 461
QY      472  ADLLOIATDYLKAHYGVDSKSNATNHLISLEAWSNDNDPOYNKOTKGAOLPIDNKLRLSL 531
      462  ADLLOVSYNFKNYKVTVDSEANALAHISILEAWSLNDNQYNEEDTNGTALSIDNSRLTS 521
QY      532  LYALTR-PLEK-DASNKNEIRSGLEPVITNSLN-NRSABGKNSERMANYIFIRADHSEVQ 588
      522  LAVLTQKQQRIDLSN-----LISESVNKEKANDATYGDITPTYSFVRADHSEVQ 571
QY      589  TVIAKIIKAQINPKTDGLTFTLDELKQAPKIYNEDMRQAKKYTQSNIPTAYALMSNKK 648
      572  TVIAKIVKEKIDTNSDGYFTLLDQLKADAFKIYNEDMAKVNKYTYHVIIPAAVALLSNNE 631
QY      649  SITRLYVGDYSDGQYMATKSPYYDAIDTLKARIKYAAGQDMKITTVYEGDKSHMDWD 708
      632  SVPRVYVGLYTDGQYMAKSPYYDAITATMLQGRYAYVSGGQSEBHVKNNGQ----- 686
```

```
QY      709  YTGVLTSVRYGTGANBATD-QGSEATKTKQGMAYVITSNNPSLKLNQNDKVIWNMGAAHKQ 767
      687  ---ILSSVRYGQDLMBSADDTQGTDLSTGLVTLVSNNDPNLDLG-GBSLTVNMGRAHANQ 742
QY      768  EYRPLLLTKDGLTSTYSDAAAKSLYRKNTDKELVFDASDIQGYLNPQVSGYLAVWVPV 827
      743  AYRPLILGTGQVQSYLKSDST-NIVKYTDANGNLTFADDIRGYSTVDMSGYLAVWVPV 801
QY      828  GASDNQDVRVAASNKANATGVYESSSALDSQLIYEGFSNFQDFTVKDSYTNKKIAQNV 887
      802  GAKGDQDVRVAADTNQKADGKSLKTSAAALDSQVIYEGFSNFQDPANNDATYTNKKIAENA 861
QY      888  QLFKSWGVTSEFEMAPQYVSEDSGFLDSIIQNGYAFEDRYDLAWSKNNKYGSOQDMINAV 947
      862  DPFKLGITSEFEMAPQYVSATDGSFLDSIIQNGYAFSDRYDLAWSKNNKYGSKDDLALANAL 921
QY      948  KALHKSGIQVIADWVDPQIYNLPCKEVTATRVNDYGEYRKDSEIKNTLYAANTKSKNGK 1007
      922  KALHANGIQAIADWVDPQIYQLPGEVVTAKRTNSYGNPTFDDAYINNALLYATNTKSSGSD 981
QY      1008  YQAKYGGAFSLAELAAKYPSIFENRTQISNGKKIDPSEKITAWKAKYFNGTNILGRGVGYVL 1067
      982  YQAYGGAFLDELKAKYPMFTVMNLISTGKPIDPSTIKIQWEAKYFNGTNVZLGAGYVL 1041
QY      1068  KDNASDKYFELKGNQTYLPKQMT-NKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGN 1126
      1042  SDDATGKYPTVNEGDFLPASFTGDQNAKTGFYDGTGMAYYSTSGNKAVNSFIYEG-GH 1100
QY      1127  WYFDNNGHMYGLQQL-NGEVOYFLSNGVQLRESFLENADGSKNTFGLGNRY-SNGYY 1184
      1101  YYYFDKGHMYTGSYKAEDGNDYIFLNPNGIQMRDAIYODAQGNSSYYGRTGILYKGDNMY 1160
QY      1185  SF-----DNDSKWRVFDASGVMAVGLKTINGNTQYFDQDGYQVKGAWITGSDGKKRYFDD 1239
      1161  PFVDPNNAKTVFRIPDANNVPAIGYRNMYGQTYIFDENGFQAKGQLLTDKQ-THYFDE 1219
QY      1240  GSGNMAVNRFANDKNGDWYLLNSDGIALVGVQTINGKTYTFG-QDGKQIKGKIITD-NGK 1297
      1220  DNGAMAKNKFVN-VGDDWYMDGNGNAVKGOYVNNQILYFNPETGVQVKGQFITDAQGR 1278
QY      1298  LKYFLANSSELARNIPAT-----DSQNNYYFG-SDGVAVTGS 1334
      1279  TSYDANSALKSSGFFTPNGSDWYYAENGYVYKGFQVAENQDQWYTFDQTTGKQAKGA 1338
QY      1335  QTIAGKGLYFASD-GKQVKSFTV-YNGKVHYTHADSGELQVNRREADKGNWYLLDSNG 1392
      1339  AKVDGRDLTFNPDSPGVQVKGDFATDESNTSFTHGNDGDKVGVGFTTGGNNAWYYADNNG 1398
QY      1393  EALTGSQRI-----NDQRVFF-TREGKQVKGDVAYD 1422
      1399  NLVKGFEIDGKWHFDEVYTGQAKGAALVNGQQLYFDVDVSGIYQVKGDFVTD 1450

RESULT 8
US-10-797-821-38
; Sequence 38, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
```

```
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 38
; LENGTH: 1554
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-797-821-38

Query Match      48.4%; Score 3615; DB 5; Length 1554;
Best Local Similarity 49.9%; Pred. No. 3.5e-179;
Matches 745; Conservative 217; Mismatches 418; Indels 112; Gaps 32;

QY 1 MEKRRYKMHKVKHVVAVAS-GLITL---GTTTLGSSVAETBQQTSDKVVTQKSED 56
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MEKLLHYKLHKVHVVAVASIGLVSLVGAGTVAEDKVANDTTAQTAVGDTGQ--- 57
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 57 DXAASESSQTDAPKT---KQAQTEQTQACS-QAANVADTSTSTKETPSONITTOANSDD 111
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 58 DQATTNDANTNTTDTDTAQSSANTNQDQAGSDQSNNDQAKQDT-----ANTDR 106
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 112 KTVNTKSEAEQTSERTKQSEEAQTASSQALTOAKAELTKORQTAQAQENKNPVDLAAI 171
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 107 NQADNSQTNQATDQATSPATDGTSGVRRDAAANVATA-ADQEGQTAPSEQEKSAAL-SL 164
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 172 PNVKQIDGKYIYIGSDGQPKNPALTVNNKVLVYFDKNTGALTDTSOYQFKQGLTKLNNDY 231
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 165 DNVKLDIGKYIVQADGSKYKPAITVNGOMLVYFSDTGALSTSTYSFSQGTNLVDDP 224
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 232 TPNHQIVNFEQTSLETIDNYTADSWYRPKDLKNGKTTWASSEDLRLPLMSWNPDKOT 291
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 225 SSHNKAYDSTAKSFELVNGVLTANSWYRPAGILRNQGTWEASNENDLRPLMSWNPDKOT 284
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 292 QYALYNNMQOGLTGENTADSSQESLNLAAGTVQVKIETKLSQTOQTOWLRDIINSFV 351
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 285 QVAYVNNMKYLSANETVNTNSQVLDNKEAQSIQTKIEQKITSNSTOWLRTAMEAFV 344
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 352 KTOPMNSQTESDTSAGEKDHLOGGALLYSNSDKTAYANSYRLNRTPTSTGTGPKYFE 411
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 345 AAQPKWNMSTE---NFNKGDLHLOGGALLYTNSDLTPWNSDYLLNRTPTTQDGTCKYFT 401
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 412 DNSSGQYDFLLANDIDNSNPVQAEQLNWLHLYLMNYGSIIVANDPEANFDGVRVDAVDN 471
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 402 EGEGGYEFLNSNDVNSNPVQAEQLNQLHLYLMNWGDIWVGDKDANFDGVRVDAVDN 461
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 472 ADLQIQTASLYLKAHYGVDKSEKNAIHNLSITLEAWSNDNDPQYNKDTKGAOLPIDNKLRLSL 531
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 462 ADLLQVSYNYFKDNYKVTSEANALAHISILEAWSLNDNQYEDTTNGTALSIDNSRLTS 521
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 532 LVALTR-PLEK-DASKNKEIRSLGEPVITNSLN-NRSAGEKNSERMANVIFIRAHDSVQ 588
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 522 LAVLTQPGQRIDLSN-----LISESVNKERANDTAYGDTIPTYSFVRAHDSVQ 571
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 589 TVIAKIIKAQINPKTDLFTLDELKQAFKIYNEDNRQAKKYTQSNITPAYALMLSNKD 648
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 572 TVIAKIVKEKIDTNSDGYFTLQDLKADPKIYNEDMAKVNKTTHYNIIPAVALLSNME 631
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 649 SITRLYVGDYSDGOYMATKSPYDAIDTLKARIKAAAGQDMKITVEGDKSHWDW 708
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 632 SVPRVYVGLYTDGGYMAKSPYDAIATMLQGRJAYVSGGSEEVHKVNGNQ----- 686
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 709 YTGVLTSVRYGTGANEATD-QQSEAKTKQMAVITSNPNSLKNQNDKVIIVNGAHKNQ 767
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 687 ---ILSSVRYGQDLMASDQTQDLSRTSGLVTLVSNDPNLDLG-GDSLTVNNGRAHAQ 742
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 768 EYRPLLLTKOGLTSTYSDAAAKSLYRKTNKDKGELVFDASDIQGYLNPQVSGYLAVVPV 827
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 743 AYRPLILGTGQVQSYLKDSDT-NIVKYTDANGNLFTTADDIKGYSTVDMSGYLAVVPV 801
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 828 GASDNDQVVAASNKANATQOVYESSSALDSQLIYEGPSNFQDPVTKDSYTNKKIAQNV 887
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
Db 802 GAKCQDQVRVAADTNQKADGKSLKTSAAALDSQVIYEGPSNFQDPANNADADYTNKKIAENA 861
QY 888 QLFKSWGVTSEFMAPOYVSSEDSPLDSIIQNGYAFEDRYDLAMSKNNKYGSQQDMINAV 947
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 862 DFFKGLGITSFEMAPQYVSATDGSFLDSIIQNGYAFSDRYDLAMSKNNKYGSKDDLANAL 921
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 948 KALHKSQIQTADWYDQIYNLPGKEVVTATATVNDYGEYRKDSEIKETLYLAANTKSNKGD 1007
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 922 KALHANGIQAIADWYDQIYQLPGBEVVTAKRTNSYGNPTFPDAYINNALLYATNTKSSGSD 981
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1008 YOAKYGGGFLSELAAKYPSIFNRTQISNGKKIDPSEKITAWKAKYFNGTNIILGRGVGYVL 1067
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 982 YOAKYGGGFLDELKAKYFDMFTVNMISTGKPIDPSTKIKQWEAKYFNGTNIILGRGVGYVL 1041
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1068 KDNASDKYFELKGNQTYLTPKQMT-NKEASTSGFVNDGNGMTFTYSTGYQAKNSFVQDAKGN 1126
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1042 SDDATGKYFTVNEGDFLPASFTGQNAKTGPYDGTGMAYYSTSGNKAVNSFIYEG-GH 1100
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1127 WTYFDNNGHMYVGLQOL-NGEVOYFSLNGVQIRESFLENADGSKNYFGLHGNRY-SNGYY 1184
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1101 YTYFDKDGHHVTSYKAEDGNDYFLPNGIQMRDAIYDQAQNSYYIGRTGILYKGDNMY 1160
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1185 SF-----DNDSKWRVFDASGVMAVGLKTINGNTQYFDQDGYQVKGAWITGSDGKKRYFDD 1239
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1161 PVDENNANKTVFRYFDANNVMAIGYRNMYGQTYTFDENGFOAKQQLLTDGK-THYFDE 1219
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1240 GSGNMAVNPANDKNGDWTYLLNSDGIALVGVQTINGKTYVFG-QDGKQIKGKIITD-NGK 1297
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1220 DNGAMAKNKFVN-VGDDWYMDGNGNAVKGQYVNNQILYFNPETGVQVKGQFITDAQGR 1278
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1298 LKYLFLANSSELARNIPAT-----DSQNNWYYFG-SDGVAVTGS 1334
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1279 TSYDANSALGKSSGFTPNGSDWYEAENGYYVYKGFQKVAENQDQWYFDDTTGKQAKGA 1338
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1335 QTIAGKLYFASD-GKQVKGSEFT-VNGKVHYVHADSGELQVNRPEADKQGNWYYLDSNG 1392
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1339 AKVDERDLVFNPDGQVQVKGDFATDESNGTSFYHGDNGDKVVGFFTTGNNAYYADNG 1398
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1393 EALTGSQRI-----NDQRVFF-TREGKQVKGDVAYD 1422
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1399 NLVKGQFQEDGKWHFDEVTTGQAKGAALVNGQQLYFDVDSGIQVKGDFVTD 1450
```

RESULT 9

```
US-09-740-274-4
; Sequence 4, Application US/09740274
; Patent No. US20020031826A1
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
```

US-09-740-274-4

Query Match 48.1%; Score 3591.5; DB 3; Length 1375;
Best Local Similarity 51.9%; Pred. No. 5e-178;
Matches 721; Conservative 215; Mismatches 350; Indels 103; Gaps 22;

QY 1 METKRYKHKVKKHVVAVASGLITLGTTLTGSSVSASAEQTSQTSKVVTKSDDKAA 60
DB 1 MEKKVRFKLRKVKRWVTVSIASAVTL-TSLSGSLVKADS-----TDDRQQA 47
QY 61 SESSQTDAPKTKQAOTEQTAQSOANVADTSTSIKETPSONITTOANSDDKTVTNTKSE 120
DB 48 VTESQASLVTSEA-AKETLTATDTSTATSQPTATVTDNVSTTNQS-----TNTTAN 101
QY 121 EAQTSERTKQSEEAQTASSQALTOAKA--ELT----- 152
DB 102 TANFVVKPTTTSQAKTDSKDIITTSKAVNRLTATGKFPANNNTAHPKVTVDKIVPIK 161
QY 153 -----KQRTAAQENKNPVDLAAIPNVK---QIDGKYIYIGSDGQPKNFALTVNNKVL 203
DB 162 PKIGKLIKQSSLSQD-----DIAALGNVKNIRKVGKYYIKEDGTLQKNYALNINGTFF 216
QY 204 YFDKNTGALTDTSOYQFQKGLTKLND---YTPHNOIVNFENTSLETIDNVYTADSWR 259
DB 217 PFDE-TGALSNNTLPSKKNIT--NNDNTNSFAQYNQVYSTDVANFEHVDHYLTABSWR 273
QY 260 PKDILKNGKWTATASSEDRLPLMSWPKQTOIAYLNTMNOQGLGTGENYTADSSQESL 319
DB 274 PKVILKDGKWTQSTKDFRPLMTWMPDQETQRYVNYMNAQ-LGIHQTYNTATSPQLQ 332
QY 320 NLAAQTVQVKIETKISQTOQTOWLRDIINSFVKTPQNNWSQTESDTSAGEKHOHLOGALL 379
DB 333 NLAAQTIQTKIEKITAENKTNLWLRQTIISAFVKTSQAMNSDSEKPPD----DHLQKCALL 388
QY 380 YSNSDK-TAVANSYKLLNRTPTSGTK--PKYFEDNSSGGYDFELLANDIDNSNPVQAE 436
DB 389 YSNSKLTQAMSNVRLNRTPTNQTKDPRDRTTIGGYEFLANDVDSNPVQAE 448
QY 437 QLNWLHLYMNGYSIVANDPEANPDGVRVDVANDVNADLLQIASDYLKAHGYVDKSEKNAI 496
DB 449 QLNWLHFLMNFNGNIYANDPDANFDSIRVDVANDVNADLLQIAGDYLKAAKGTHKNDKAA 508
QY 497 NHLSTLEAWSNDPQNVKDTGAQPIIDNKLRLSLYALTREPKDASKNRISLEPV 556
DB 509 DHLSTLEAWSYNDTPYLDHGDGNMNMNRLSLSLYSLAKPLNQ-----RSGMDEL 560
QY 557 ITNSLNNSAEGKNSERMANFYIFRAHDSVQTVIAKIIKAOINPKTDGLTFLDELQA 616
DB 561 ITNSLVNRTDDNAETAAPVPSYFIRAHDSVQDLIRNIIIRTPNPVVGYSPTTEIKKA 620
QY 617 FKIVNEDMRQAKKYTQSNIPYALMLSNKDSITRLYYGDMVSDDGQYMATKSPYDAI 676
DB 621 FEIYNKDLLATEKKYTHNTALSYALLTNKSVPRVYGGDMFTDDGQYMAHKTINYEAI 680
QY 677 DTLKARIKIAAGGQDMKITIYVEGDKSHMDWDYTGVLTSVRYGTGANEATDQGSBATKQ 736
DB 681 ETLKARIKYVSGGQMRNQGVNSE-----IITSVRYGKALKATDGTDRTRFS 731
QY 737 GNAVITSNPSLKLNDQKVIYVNGAAHKNQBYRPLLLTTKDLGTSYTSDDAAKSLYRKT 796
DB 732 GVAVTEGNPSLRLKASDRVVVNMGAHKNQBYRPLLLTTDNGIKAYHSDQBEAAGLVRYT 791
QY 797 NDKGBLVPDASDIQGYLNPQVSGYLAVVVPVUCASDNQDVRVAASNKANATQGVYSSAL 856
DB 792 NDRGELIITADIKGYANPQVSGYLGWVPVPGAAADQDVRVAASTAPSTDGKSVHQNAL 851
QY 857 DSQLIYEGSNFQDFVTTKSDYTNKKIAQNVLQFKSWGVTSPFEMAPQYVSSDGSFLDSI 916
DB 852 DSRVMFEGSNFQAATKKEEYTNVIAKNVDKFAEWGVTDEMAPQYVSSDGSFLDSV 911
QY 917 IONGVAFEDRYDLAMSKNKYGSQODMNAVVALHKSIGQVITADVWPQIYNLPKEVVT 976
DB 912 IONGVAFEDRYDLGISKPNKGTADDLVKAIKALHKSIGIKVMADVWPQVMTALPEKVVT 971

QY 977 ATRVNDYGEYRKDSEIKNTLYAAANTKSNKDYQAKYGGAFSLSEAAKYPGSIENRTOISNG 1036
DB 972 ATRVDKYGTPVAGSQIKNTLYVVDGKSSGKQQAQYGGAFLEELQAKYPFELFARKQISTG 1031
QY 1037 KKIDPSEKITAWAKAFNGTNIILGRGVGYVLKNASDKYPFELKGNQTYLPKQMTN----- 1091
DB 1032 VPMDPSVKIKQMSAKYFNGTNIILGRGAGYVLKQATNTYPSLSVSDNTFLPKSLVNPNGHT 1091
QY 1092 KEASTQFVNDGNGMTFYSTGYOAKNSFVQDAGKNVYFDNNGHMYVGLQOLNCEVOYEL 1151
DB 1092 SSVTGLVFDGKGYVYISTSGNQAKNAFI-SLGNMYIFDNNGYMTVGTGAISINGANYFEL 1150
QY 1152 SNGVQLRESFLENADGSKNYFGLHGNRYSGYYSFNDNSKWRYPFASGVMAVGLKTINGN 1211
DB 1151 SNGIQLRALNIYDNGNKVLSYNGNDGRRYGYLFP--GQQRIFYQ-NGINAVGLTRVGA 1207
QY 1212 TQYFDQDQGVQKAMWITGSDGKRYFPDGSNNMAVNRFANDKNGDWYLLNSDGLIALVGVO 1271
DB 1208 VOYFDASGFQAKQFITTADGKLRYFDRDSGQISNRFVRNSKGEWFLFDHNGVAVTGTV 1267
QY 1272 TINGKTYYPGQDGKQIKGIITD-NGKLXYFLANSGLARNIATDSQNNWYFSGSDVA 1330
DB 1268 TTNGQRLYFPKNGVQAKGEFIRDANGLYRYDPNSGNEVRNFRVRNSKGEWFLFDHNGIA 1327
QY 1331 VTGSQTIAG 1339
DB 1328 VTGARVVG 1336

RESULT 10

US-10-383-930-35
; Sequence 35, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 35
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-35

Query Match 48.1%; Score 3591.5; DB 4; Length 1375;
Best Local Similarity 51.9%; Pred. No. 5e-178;
Matches 721; Conservative 215; Mismatches 350; Indels 103; Gaps 22;

QY 1 METKRYKHKVKKHVVAVASGLITLGTTLTGSSVSASAEQTSQTSKVVTKSDDKAA 60
DB 1 MEKKVRFKLRKVKRWVTVSIASAVTL-TSLSGSLVKADS-----TDDRQQA 47
QY 61 SESSQTDAPKTKQAOTEQTAQSOANVADTSTSIKETPSONITTOANSDDKTVTNTKSE 120
DB 48 VTESQASLVTSEA-AKETLTATDTSTATSQPTATVTDNVSTTNQS-----TNTTAN 101
QY 121 EAQTSERTKQSEEAQTASSQALTOAKA--ELT----- 152
DB 102 TANFVVKPTTTSQAKTDSKDIITTSKAVNRLTATGKFPANNNTAHPKVTVDKIVPIK 161
QY 153 -----KQRTAAQENKNPVDLAAIPNVK---QIDGKYIYIGSDGQPKNFALTVNNKVL 203
DB 162 PKIGKLIKQSSLSQD-----DIAALGNVKNIRKVGKYYIKEDGTLQKNYALNINGTFF 216

Db 389 YSNNSKLTQANSNRYLRNTPNTQTKGKDPRTYADRTIGGYEFLLANDVNSNPPVQAE 448
Qy 437 QLANWLHLYMNGSVIVANDPEANFQGVVRVDAVNVNADLLOQIASDYLKAHYGVDKSEKNAI 496
Db 449 QLANWLHLYMNGSVIVANDPEANFQGVVRVDAVNVNADLLOQIASDYLKAHYGVDKSEKNAI 508
Qy 497 NLSILEAWSDNDPOYNKDTKGAQLPDKNLRLSLLYALTRPLEKDNKNEIRSGLEPV 556
Db 509 DLSILEAWSYNDTPYLLHDPGDNMINNDNRLSLLYSLAKPLNQ-----RSGNPL 560
Qy 557 ITNSLNRSAGKNSERMANVIFIRAHDSVQTVIAKIIKAIKPNKTDGLTFLDELKQA 616
Db 561 ITNSLVNRTDNETAATAVPYSFIRAHDSVQDLIRNIIIRTEINPNVVGYSFTTEIKKA 620
Qy 617 FKINEDMRQAKKYKTYTOSNPTAYALMLSNKDSITRLYYGDMYSDDDQGYMATKSPYYDAI 676
Db 621 FEIYNKDLATEKKYTHYNTALSVALLLTNKSSVPRVYVYVGMFTDDCGYMAHKTINYEAI 680
Qy 677 DTLKARIKAAAGQDMKITYVEGDKSHMDWDYGVLTSTVRYGTGANEATDQGEATKQ 736
Db 681 ETLKARIKIVSGQAMRNQOVGNSE-----IITSVRYGKALKATDGTDRTRTS 731
Qy 737 GMVITSNPSLKLNDKVIYVNGAAHKNQOYRPLLLTKDGLTSYTSDAAKSLYRKT 796
Db 732 GVAVIEGNNPSLRKASDRVVVNGAAHKNQOYRPLLLTDNGIKAYHSDQEAAGLVRYT 791
Qy 797 NDKGELVFDASDIQGYLNPQVSGYLAVVWVPVGASDNQDVRVAASNKANATGVYVESSSAL 856
Db 792 NDRGELIFTAADIKGYANPQVSGYLVWVPVGAADQDVRVAASTPSTDGKSVHQAAL 851
Qy 857 DSQIIVEGFSNFQDFTVKDSQDYNKKIAQNVQLFKSGWVTSFEMAPQYVSEBSGFSLDSI 916
Db 852 DSRVMEFEGFSNFQAFATKKEEYVNVVIAKNVDFAEWGVTDFFEMAPQYVSSTDGFSFLDSV 911
Qy 917 IQNGYAFEDRYDLAMSNKNKYGSQDMINAVKALHKSIGIOVADWVPDQYLYNLPGEVYT 976
Db 912 IQNGYAFEDRYDLGISKPNKIGTADLLVKAIKALHKSIGIKWADWVPDQYLYNLPGEVYT 971
Qy 977 ATRVNDYGEYRDKSEIKNTLYAANTKSNKGDYQAKYGGAPLSLAAKYPSIFNRTOISNG 1036
Db 972 ATRVDKYGTPVAGSQIKNTLYVVDGKSGKQDQAKYGGAPLSLAAKYPSIFNRTOISNG 1031
Qy 1037 KKIDPEKITAWEKAKYFNGNIIILGRGVYVYKDNASKYFELKGNQTYLPKQMTN----- 1091
Db 1032 VPMDPSVKIKQWSAKYFNGNIIILGRGVYVYKDNASKYFELKGNQTYLPKQMTN----- 1091
Qy 1092 KEASTGTVNDGNGMTFVSTGYOAKNSFVQDAGKNMYFDNNGHMYVGLQOLNGEVQYFL 1151
Db 1092 SSVTVGLVFDGKGYVYTSNGQAKNAFI-SLGNMYFFDNNGYMYTGAQINGANYFL 1150
Qy 1152 SNGVQLRESFLENADGSKNYFGLGNRYSGYYSFNDNSKWRVFDASGVMAVGLKTINGN 1211
Db 1151 SNGIQLRNAIYDNGKVLVSYGNDGRYENGYYLF--GQWRVYFQ-NGIMAVGLTRVHGA 1207
Qy 1212 TVQFDQDGYQVKAWTISDGKRYFDDGSGNMAVNRFANDKNGDYYLNSDGIALLVGVQ 1271
Db 1208 VOYFDSAGFQAKGQFITTADGKLYRFDSDSGNQISNRFRVNSKGEWFLFDHNGVAVTGTV 1267
Qy 1272 TINGKTYFQDQKQIKGLIITD-NGKLYFLANSSELARNIPATDSQNNWYFSGDGA 1330
Db 1268 TFGQRLYFPKPVQAKGEFIRDANGLYRYPDPSNGNEVRNFRVNSKGEWFLFDHNGIA 1327
Qy 1331 VTGSQTIAG 1339
Db 1328 VTGARVNG 1336

RESULT 12

US-10-383-930-37
; Sequence 37, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J

; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 37
; LENGTH: 1590
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-383-930-37

Query Match 47.9%; Score 3575; DB 4; Length 1590;
Best Local Similarity 48.4%; Pred. No. 4,3e-177;
Matches 719; Conservative 240; Mismatches 408; Indels 118; Gaps 26;

Qy 1 METKRYKMKHKVKKHWTVAVASGLITLTGTTTIGSSVSAETEQTSDKVVTQKSEDDKAA 60
Db 1 MEKNVRFKHKVKKRWVTLISVASA--TMLASALGASVAS-----ADDTAS 44

Qy 61 SESSQTDAPKTKQAQTEQTQAQSQANVADTSTSTIKETPSQNTTQANSDDKTVTNTKSE 120
Db 45 DDSNQA---VVTGDTNNQATDQTSIAATATS-EQASATDADTQASAEQTQGTAST 100

Qy 121 EAQTSERTKQSEEAQ---TTASSQALTOAKAELTKOROTAAQENKNPVDLAAIPNVKQI 177
Db 101 D--TAAQTITNANEAKWPTENENQGTDEMLAEAKNVAT-AESDSTPSDLAKMSNVKQV 157

Qy 178 DGKYYIGSGQPKNFALTNNKVLVFDKNTGALTDTSQYQFKQGLTKLND---YTPH 234
Db 158 DGKYYIYDQGNVKNFVSVGDKIYYFDE-TGAYKDTSKVDADKSSASVQNTATFAAN 216

Qy 235 NQIVNPENTSLKETIDNYVTADSWYRPKDLKNGKWTATSSSEDLRLPLLSWPDQTOIA 294
Db 217 NRAYSTSAKNFEAVDNYLTADSWYRPKSLKDGKWTESGKDDFRPLLMAMWPDTEKRN 276

Qy 295 YLYNMNQGLGTCENTADSSQESLNLAQTVQVKIETKISQTOQOTOWLRDIINSFVKQ 354
Db 277 YVYMNKV-VGIDKTYTAETSQADLTAAAEVLQARIEQKITSNNNTKWLREALSFAVKQ 335

Qy 355 PNWNSQTESDTSAGEKDHLOGGALLYSN-SDKTAYANSYRLNRLNRPSTQGT-----KPKY 409
Db 336 PQMGSEKPYD-----DHLQNGALLFDNQDTDLTPDQSNRYRLNRLNRPSTQGTSLDSRFTY 391

Qy 410 FEDNSGGYDFLLANDIDNSNPVQAEQNLNWLHLYMNGSVIVANDPEANPDGVRVDAVDN 469
Db 392 NPNDPLGGYDFLLANDVDNSNPVQAEQNLNWLHLYMNGSVIVANDADAFDSIRVDAVDN 451

Qy 470 VNADLLOIASDYLKAHYGVDKSEKNAIHLISLEAWSNDNDPOYNKDTKGAQLPIDNKLRL 529
Db 452 VDADLLOISSDYLKAAAGIDKKNKANNHVSIVEAWSNDNDTPYLHDDGDNLMNMDNKFRL 511

Qy 530 SLLYALTRPLEKDNKNEIRSGLEPVTITNSLNRSAGKNSERMANVIFIRAHDSVQTV 589
Db 512 SMLWSLAKPLDK-----RSLGNPLIHSNLDREVDREVEVTPSVSFAFAHDSVQD 563

Qy 590 VTAKIIRKAIQINPKTDGLTFLDELKQAQKIYNEDMRQAKKYTOSNPTAYALMLSNKDS 649
Db 564 IIRDIIRKAEINPNSFGYSFTQEIQAQFKIYNEDLKKTKKTHYVNPVLSYLLLTNKG 623

Qy 650 ITRLYYGDMSYDDGYMATKSPYYDAITLTKARIKAAAGQDMKITYVEGDKSHMDWDY 709
Db 624 IPRVYVYVGMFTDDGQYMAKNTVNYDAIESLLKARMKYVSGGQAMQ-NYQINGE----- 676

Qy 710 TGVLTSTVRYGTGANEATDQGEATKQGMVAVITSNPSLKLNDKVIYVNGAAHKNQOY 769
Db 677 --ILTSTVRYGKALKQSKDGDATRTTSGVGVVGMGNQPNFSLD-GKVVALNMGAAHANQOY 733

QY 770 RPLLLTTKDLGTSYSDAAK--SLYRKTNKDELVEFDASDIQGYLNPQVSGYLAVWVPV 827
Db 734 RALMVSTKDGVTATYATDADASKAGLVKRTDENGSLYFLNDDLKGVANPQVSGFLQVWVPV 793
QY 828 GASDNQDVRVAASNKANATGVQVSESSALDSQLIYEGFNFQDFVTKDSQDYNKKIAQNV 887
Db 794 GAADDQDIRVAASDTASTDGKSLHQDAAMDSRVMPFSGFNFQSFATKESEYTNVVIANNV 853
QY 888 QLFKSGVTSFEMAPQVYSEDSGLDSIQNGYAFEDRYDLAMSNNKKNYKSGQDMINAV 947
Db 854 DKFVSGITDFEMAPQVYSSDQFLDSVQNGYAFTRDYDLGMSKANKYGTADQLVKA 913
QY 948 KALHKSGIOVIADWPDQIYNLPCKEYVATRVNDYGEYRKDSEIKNTLYAANTKSKNGD 1007
Db 914 KALHAKGLKVMADWPDQYVTPFKQEVVTRTDFKFKPIAGSQIHNSLYVTDYTKSSGDD 973
QY 1008 YQAKYGGAFSELAAYGPIFNRTQISNGKKIDPSEKITAWKAKYFNGTNNILGRGVGYVL 1067
Db 974 YQAKYGGAFDELKEKYPFLTKQIISTGQIDPSVKIKQWSAKYFNGSNILGRGADYVL 1033
QY 1068 KONASDKYFELKGNQTYLPRQMTNKEASTGFVNDGNGMTFYST--SGYQAKNSFVQDAKN 1126
Db 1034 SDQVSNKYFNVASDITLFLPSSLLGKVVEGIRYDGGYIYNSSATGDQVKASPIEA-GN 1092
QY 1127 WYFPDNNHMYGLQOLNGEVQYFLSNGVQLRESFLENADGSKNYFCHLGNYSNGYYSF 1186
Db 1093 LXYFPGKDGVMYVTAQTINGANYFPFLENGTALRNTIYTDAGNSHYVANDGKRYENGYYQF 1152
QY 1187 DNDKRWYFDASGVMAVGLKTNGNTQYFDQDGYQVKGAWITGSDGKKRYFDDGSGNMAV 1246
Db 1153 GND--WRYF-KDGNMAVGLTVDGNVQYFDKQGVQAKDLIVTRDQKVYFYFOHNGNAAT 1209
QY 1247 NRPANDKNGDWYILNSDGLALGVQTINGKTYFYFGDQKQIKGIIT-DNGKLKLYFLANS 1305
Db 1210 NTFIADKTHWYILGKDGVAVTGAQTVGKQLYFEANGQVKGDFVTSDEGKLYFYDVIDS 1269
QY 1306 GELARNIFATDSQNNWYFGSDGVATGSGQTAGKLYFASDGKQVKGDFVT--YNGKVHY 1364
Db 1270 GDMWTDFTIEDKAGNWFYLGKOGAAVTGAQTIRGQKLYFKANGQQVKGDIVRGTDGKIRY 1329
QY 1365 YHADSLEQVNRPEADKDG----- 1383
Db 1330 YDAKSGEVPENKTVKAADGKTYVIGNDGVAVDPSVVGQTFKDKDASGALRFPYNLKQIIVTG 1389
QY 1384 -----NWYLDNSGREALTGSQRINDQRFVFTREGKQVKGDV 1419
Db 1390 SGWYETANHDWYIQS-GRALTGEQTINGQHLFYFKEDGHQVKGQL 1433

RESULT 13

US-10-797-821-37
; Sequence 37, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45

; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 37
; LENGTH: 1590
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-797-821-37

Query Match 47.9%; Score 3575; DB 5; Length 1590;
Best Local Similarity 48.4%; Pred. No. 4.3e-177;
Matches 719; Conservative 240; Mismatches 408; Indels 118; Gaps 26;

QY 1 METKERYKMHKKYKHWTVAVASGLITLTGTTTLGSSVSASAEETQOTSQDKVVTQKSEDDKAA 60
Db 1 MEKNVRFMKHKYKHWTVLSVASA--TMLASALGASVAS-----ADTDITAS 44
QY 61 SESSQTDAPKTKQAQTEQTAQSQANVADTSTSIKTETPSQNTITQANSDDKTVTWTKSE 120
Db 45 DDSNQA--VVTGDQTTNNQATDQTSIAATATS-EQSASTDAATDQASAAEQTQGTAST 100
QY 121 EAQTSEERTKQSEBAQ---TTASSQALTOAKAELTKQRTAAQENKNPVDLAAIPNVKQI 177
Db 101 D--TAQTNTTNAEAKWPTENENQGFTEMLAEAKNVAT-AESDSIPSDLAKMSNVKQV 157
QY 178 DGKYYIIGSDGQPKKNFALTNNKVLYPDKNTGALTDTTSOYQFKOGLTKLNND---YTPH 234
Db 158 DGKYYIYDQGNVKNFVSGVKIYYFDE-TGAYKDTSKVDADKSSVASQNTIFAAN 216
QY 235 NQIVNFENTSLEITDNYVTADSWYRPRKOILKNGKTWTASSESDLRPLLLMSWMPDKQTOIA 294
Db 217 NRAYSTSAKNFEAVDNYLTADSWYRPKSILKDGKTWTESGKDDFRPLLLMAWMPDTETKRN 276
QY 295 YLYNNVQOGLTGNGENTADSSQESNLAAQTVOVKIETKISOTOQTOWLRLDINSFVKQ 354
Db 277 YVYNNKVV-VGIDKTYTAETSQADLTAAAEVQARIEQKITSENNTKWLREASIAFAVKTQ 335
QY 355 PNWNSQTESDTSAGEKHLQGGALLYSN--SDKTAVANSDYRLNLRTPSTQSG---KPKY 409
Db 336 POWNGESEKPYD----DHLQNGALLFDNQTDLTPTDQSNYRLNLRTPNQTSGLDSRFFY 391
QY 410 FEDNSSGGYDFLLANDIDNSNFPVQAEQLNWLHLYLMNYGSIIVANDPEANFDGVRVDAVDN 469
Db 392 NFNDPLGGYDFLLANDVDNSNFPVQAEQLNWLHLYLNFGSIVANDADANFDSIRVDAVDN 451
QY 470 VNADLLQIASDYLYKHYGVDKSEKAINHLSILEAWSNDNDPQYNKDTKGAQPLIDNKLXL 529
Db 452 VDADLLQIASSDYLYKAAYGIDKNNKANNHVSIVEAWSNDNTPYLDHDDGDNLMNMKNFRL 511
QY 530 SLLYALTRPLEKDAASKNEIRSGLEPVTINSNNRSAGCKNRMANYIPIRAHDSQVOT 589
Db 512 SMLWSLAKPLDK-----RSLGPLIHNSLVDREVDDREVETVPSYSFARAHDSQVQD 563
QY 590 VIAKIIKAQINPKTDGLTFLDELKQAFKIYNEDMRQAKKQYTSQSNIPTAYALMLSNKDS 649
Db 564 IIRDIKABINPNSFGYSFTQEEIEQAFKIYNEDLKKTKDKKYTHYNVPLSYTLTLTNKGS 623
QY 650 ITRLYYGDMSDQGYMATKSPYYDAIDTLKAKIYKAAAGGQDKMITYVEGDKSHMDWDY 709
Db 624 IPRVYVYGDMTDGGYTMANKTYNYDAIESLLKARMKYVSGGQAMQ-NYQINGE----- 676
QY 710 TGLTVSVRYGTGANEATDQSEATKTQGMNAVITSSNPNLSLKNQNDKVIYVMGAHKNQY 769
Db 677 --ILTSVRYGKALKQSDKGDATTTRISGVGVVNGNQPNFSLD-GKVVALNMGAAHANQY 733
QY 770 RPLLLTTKDLGTSYTSDDAAK--SLYRKTNKDELVEFDASDIQGYLNPQVSGYLAVWVPV 827
Db 734 RALMVSTKDGVTATYATDADASKAGLVKRTDENGSLYFLNDDLKGVANPQVSGFLQVWVPV 793
QY 828 GASDNQDVRVAASNKANATGVQVSESSALDSQLIYEGFNFQDFVTKDSQDYNKKIAQNV 887
Db 794 GAADDQDIRVAASDTASTDGKSLHQDAAMDSRVMPFSGFNFQSFATKESEYTNVVIANNV 853
QY 888 QLFKSGVTSFEMAPQVYSEDSGLDSIQNGYAFEDRYDLAMSNNKKNYKSGQDMINAV 947

Db	1239	SKVVKFRYFTN	EGVMAKVTVT	VDGFTQY	FGEDGFQAKDKLVT-FKGKTYTYFDAHTNGXIK	12997
Qy	1247	NRFANDXNGDWY	YLNSDGIALV	GVQTINGKTYTY	FGDQKGKIKITD-NGKLKYFLANS	1305
Db	1298	DTWRN-INGK	WYFDANGVAAT	GAQVINGKLY	FNEDGSQVKGVGKVNADGTYSYKEGF	1356
Qy	1306	GELARNIFATDS	QNNWYTFGSDG	VAVTGSQTIAGK	LYFASDGSQVKGVSFV-TYNGKVHY	1364
Db	1357	GELVTNEFF	TDGNVWYTAG	KNTVTAQVINGQ	HLYFNADGSQVKGVGKVNADGTYSK	1416
Qy	1365	YHADSGSLQVNR	FEADKGNWY	LDNSGEAL	TGSQRINDORVFFTRREGKQVGD-VAYDE	1423
Db	1417	YNASTGERLTNE	FFTTGDDNNWY	YGANGKSVTGEVK	IGDDTYTFFAKDGQVKGQTVSAGN	1476
Qy	1424	RLLLVY	1429			
Db	1477	GRISY	1482			

Search completed: February 11, 2006, 20:56:52
Job time : 134.994 secs

100

THIS PAGE BLANK (USPTO)

Db 902 INKTESIFVETE---KTFSEYANHITBEISKIKGTIFDTVNGKL---VKKVNLDTTH 955
Qy 234 HNOIVNPE---NTSLTIDNVVTADSWYRPDKILKNGKTW-----TA 272
Db 956 EVNTLNAAFIOSLIENSSESUSLSNVAMKVQVYAQLFSTGLNTTDAKVVVELVSTA 1015
Qy 273 SSSES-DLRPLMSWPKQTOIAYLNTMNOQGLGTGENYTADSSQESLNLAQAQTVQVKIE 331
Db 1016 LDETIDLPTLSEGLPIATII-----DGVSLGA-AIKELSETSDPLLRLQIEAKI- 1065
Qy 332 TKLSQTOOTOWLRDILINSFVKTPQNW-----NQTESDTSAGEKH 372
Db 1066 -GIMAVNLTTATTAITTSGLAGSFGSILLVPLAGISAGIPSLVNNELVLRDKATKVVDY 1124
Qy 373 LOGGALLYSN-----SDKTAYANSYRLLNRPPTSOTGPKPYFEDNS-----S 415
Db 1125 FKHSVLETGVTLLDDKVMQDDLLV-----SEID-----FNNSIVLGKEIWRME 1174
Qy 416 GGYDFLLANDIDN--SNP-----VVOAEQLN-----439
Db 1175 GSGGHTVTDIDHFFSAPSITYREPHLSIYDVLEVQKEELDLSKOLMVLNPNRVPFAWE 1234
Qy 440 --W---LHLYMNGSIVANDPEANFDG-----VRVDAVD 468
Db 1235 TGWTPGLRSLENDGTLLDRIDRNYEGEFYRWYFAFTADALITTLKPRYEDTNRINLDS 1294
Qy 469 NVNADILQ-IASDYLK-----AHYGVDKSEKN-AIN-HLSILEAWSNDNQYNK 514
Db 1295 NTRSFVPIITTIYIREKUSYSYFGSGGTVALPLSQYNMGINIELSESDVMI-----1346
Qy 515 DTKGAOLPIDNKLRLSLYALTRPLEKADSNKNEIRSGLEPVTINSNNRSARGKNSERM 574
Db 1347 -----IDVDNVVR-----DVTIESDKIKKGLIEGI-----LSTLSIE-----1379
Qy 575 ANYIFIRANDSEVQTVIAKIIKAQINPKTD--GLTFTL-----DELKQAFKI---619
Db 1380 ENKIIILNSHEIN-----FSGEVNGSNGFVSLTFSILEGINAIIIEVDLLSKSYKLLIS 1431
Qy 620 -----YNEDMRQAKKY-----TQSNITPTAYA-----LMLSNDKS-----649
Db 1432 GELKIILMNSHIIQKIDYIGFNSQLKNIPIYSFVDSGEKNGFINGSTKEGLFVSELDP 1491
Qy 650 ----ITRLYIGDMYSDGOYMATKSPYYDAIDTLKKARIKYAAG---GQDMKITY---VEG 700
Db 1492 VVLISKVYMDSDKSPSGY-----SNLKDVKVITKDNVAILTGYLKDIKISLSLTQD 1547
Qy 701 DKS-----HMDWYTGVLTSVRYGTGANEATDQGEATKTQGMVITSNNPSLK-----749
Db 1548 EKTIKLNSVHL--DESGVAELIKF-----MNRKSGTNTSDLSMSFLESNN--IKSIFVN 1597
Qy 750 -LNQNDKVIWNMG-----AAHKQOEYRPLLLITTKDGLTSVTSDDAAKSLY- 793
Db 1598 FLQSNIKFILDANFIISGTTSIGQFEFICDENNNIQFYFIKFWLTETNTY-----LVV 1650
Qy 794 -RKTNDKGLVFDASDIOGYNLPQVSGYLAVVPVCGASDNQDVRVAASKANATGOVYES 852
Db 1651 GNRQNWIVPNVDLDP--SDGISSTVINFSQKYL-YGID-----SCVNKVIISPNY--1699
Qy 853 SSALOSQLIYEGFSNQDFVTKDSYTNKKIAQNVLQFKSWGVTSPFMAPQVYSSBDGS- 911
Db 1700 TDEINITPVYETNTVPEVILVDANYINEKINVIN-----DLSIRVWVNSDND 1749
Qy 912 -----FLDSIIQN--GYAFEDRYDLAMSK-----933
Db 1750 FILMSTSEENKVSQVKIRVNVNFKOTLANKLSFNFSRQDQVPVSEIILSFTSPSYVEDGL 1809
Qy 934 -----NNKYGSGQDMINAVKALHKS-----GIQVIADWVD 964
Db 1810 IGYDGLVSLYNEKFPVNNFGM---MVSGLIYINOSLYYFKPPVNNLITGFTVTGVD---D 1863
Qy 965 QIYNLP---GKEVWTATRVNDYGEYRKDSEIKNTLYAANTKSGNDKYQAKYGAGFLSELA 1021

Db 1864 KYTFNPNINGGAASIGETIIDDKNYFNQSGVLQT-----GVFSTEDG 1905
Qy 1022 AKYPSIFNRRTQIS-NGKKIDPSEKITAOKAKYFNGTNILGRGVYVLKDN---ASDKYFEL 1078
Db 1906 FKYPAPANTULDENLEGEAID-----FTGKLIIDENI--YYPEDNYRGAVWEKEL 1952
Qy 1079 KGNQTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVODAKGNWYFYFDNNGHMVY 1138
Db 1953 DGEWHYFSPF-TGK-AFKGLNQIGDDKYFNSDGVMQK-GFV-SINDNKHVYFDDSGVMKV 2008
Qy 1139 GLOQLNGEVOYFLNSGVOLRESFLENADGSKNTYFGH-----LGNR-----YSNGYYSFDN 1188
Db 2009 GYTEIDGKHFFPAENG-EMOIGVENTEDGFK-YFAHHNEDLGNBEGEIEISYS-GILFNFN 2065
Qy 1189 DSKWRYFDASGVMAVGLKTINGNTQYFDQGYQVKGAWITGSGKKRYFDDGSGNMAVN- 1247
Db 2066 --KIYFPDDSFYAVGWKDL-----EDGSKYFFDEDTAEAYIGL 2102
Qy 1248 RFANDKNGDWYILNSGIALVGVQTINGKTYFYFQDGKQIKGK-----IITDNGKL- 1298
Db 2103 SLINDGQ---YYFNDGIMQVGFVTTINDKVFYFSDSGIIESGVQINDDNYFIDNIGIVQ 2159
Qy 1299 -----KYFLANSSELARNIFAT-----DSQN 1319
Db 2160 IGVFTSDGYKYF-APANTVNDNIYQVAEYSGLVRVGEDVYFYGETYTIETGTWYDMEN 2218
Qy 1320 N-----WYFSGDGVAVTGSOTIAGKLYFASDQKQVKSFV 1356
Db 2219 ESDKYFVFPETKACKGINLIDDIKYFDEKGMRTGLISFENNYYFNENG-BIQFGYI 2277
Qy 1357 TYNGKVHYHADSGELQVNRFEADKDNWY-----LDSN--GEAL--TGSORINDORVF 1407
Db 2278 NIEDKMFYGED-GVMQIGVFNT-PDGFKYFAHQNTLDENFESESINTGWLGLDEKRY 2335
Qy 1408 FTREGKQVKGDAVD 1422
Db 2336 FTDEYIAATGSIID 2350

RESULT 3

US-11-052-554A-217
; Sequence 217, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 217
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Streptococcus mutans UA159
US-11-052-554A-217

Query Match 5.0%; Score 370; DB 7; Length 726;
Best Local Similarity 27.0%; Pred. No. 2.8e-12;
Matches 110; Conservative 42; Mismatches 137; Indels 118; Gaps 10;

Qy 999 ANTKNGKDYQAKYGAFLELAAYKPSIFNRTQISNGKKIDPSEKITAOKAKYFNGTNI 1058
Db 67 AKTDSANKDPQKBTG-----SVATDAPSNMNSQSDKQNTVNEI-----108
Qy 1059 LGRGVGYVLKDNASKYFELKGNQTYLPK-OMTNKEASTGFVNDGNGMTFYSTGYQAKN 1117
Db 109 -----SSDSQQTKTDEQTDLPQNSFKQSAHVKMTTEAEKTPSHSI-----N 150

QY 213 TDTSQYQFKGLTKLNNDYTPHNOIVNFENT--SLETIDNYVTADSWYRPKDIILKNGKTW 270
Db 746 ADGSSRIEIGNTIQQDKKNAASINDILNTGFLKNNNNPIDFVSTYDIVD-FANGNAT 804
QY 271 TASSEDLRPLLLMSWPDKQTOIAY-----LNYMNOQGL 304
Db 805 TATVTHDT-----ANKTSKVYDVNVDDTTIHLTGTDNKKLGKVTYKLNKTSANG- 855
QY 305 GTGENYVTADSSQESLNLAOTVQVKIETKISQTOQTQWLRLDI----- 346
Db 856 NTATNFVNSSDEDALVNAKDIAENLNLAKIHTTKGTADTALQTFVVKVDENNADD 915
QY 347 INSEFVKTOPWNQTESDTSAGE-----KDHLOGGALLYSNSDKTAYAN-----SDYELLN 397
Db 916 ANAITVQCKNANQVNTLLTKGNGLNKTKDKNGTVTFGINTTSGLKAGKSTLNDGGLSI 975
QY 398 RPTPTS-----QTGRP--KYFEDNSSGGYDFLLANDIDNSNPVQAEQLNLHLYLMNYSIV 451
Db 976 KNPTGSEIQVAGADVKAQVNNNG-----VVGAGIDGTTIRDE-----IGFTGTNGSLD 1027
QY 452 ANDPEANFDGVRVDVNDVNNADLLQIASDYLKAHYG-----VDKSEKNA 495
Db 1028 KSKPHLSKOGINAGGKKITNIQGEIAQNSHDAVTGKIYDLKTELENKISSAKTAQNS 1087
QY 496 INHLSILEAWSN-----DPQYNKDTKGAQLPI-----DNKRLSLLYALTRPLEKDSAKN 547
Db 1088 LHEFSVADEOGNFTVSNPYSSYDTSKTSVDITFAGENGITTKV-----NKG 1134
QY 548 EIRSGLE-----PVITNSLNKRSABGKNSEMANVIFIRAHDSVQTVIAKIIKAQIN 600
Db 1135 VVRVGIDQTKGLTTPKLTGNNN-----GKI-----VIDSQNG 1168
QY 601 PKT-DGLTFTLDELKQAFKIYNDMRQAKKKTQSNIPYALMLNSKDSITRLYIGDMY 659
Db 1169 QNTITGLSNTL-----ANVTNDKGSVRTEQGNII 1198
QY 660 SDDQGYMYKSPYDAIDTLKARIKVAAGQDMKIYVEGDKSHMDWDYTGVLTSVRYG 719
Db 1199 KOEDKTRAA-----SIVDLSAGFLNQNGEAV-----DFVSTYDTVNEA 1238
QY 720 TGANEATDQSEATKQGMVITSNPVLKLNQNDKVIINMGAAHKNQEVRLPLLTQDG 779
Db 1239 DG-NATT-----AKTYDDTS-----KTSKVVDVNVDDTTIEVKDKKLGKVG-- 1279
QY 780 LYSYSDAAAKSLYRKTNB--KGEVFDASDIOGLNPQVSGYLAVWVPVGCASDNQDVRVA 838
Db 1280 TTTLTSTGTGANKFALSQATGDALVKASDIVAHLN-TLSG-----DIQTA 1324
QY 839 ---ASNKANATQVYESSSALDSOLIIEGFSNFODFTVKDSYTNKKIAQNVQLFKSWGVT 896
Db 1325 KGASQANNAGYV-----DADGNKVIYD-----STDNKY----- 1353
QY 897 SFEMAPQYVSSSEGSFLDSIIQNGYAFEDRYDLAMSKNNKYGQQDMINAVKALHKSIGQ 956
Db 1354 -----YQAKNDGT-----VDKTEVAKDKLV 1374
QY 957 VIADWVFDQIYNLPGEVVVATRVNDYGBYRKDSKIKNTLYAANTKSKNGDYQAKYGGAF 1016
Db 1375 AQAQTPDGLTQWNVKSVINKQVNDAN--KKQGINEDNAFVKGLEKASDNKTKQAATV 1432
QY 1017 LSEL--AAKYPISFNRTQISNGKKIDPSEKI TAWKAKYFNGTNI LGRGVGYVLKDNASDK 1074
Db 1433 VGLDNAAVQTPLTTFAGDTGTAKKL--GETLTI-KGGQTDNKLTDNNIGV-----AGTD 1485
QY 1075 YFELKGNQTYLPKQWNTKEA--STGFPVNDGNGMTFTYSTGYQAKNSFVQDAGKNWYFYN 1132
Db 1486 GFTVK-----LAKDLTNLNSVAGGTTKIDDKGVSFVDSGQAKANTPVLISANG-----LDL 1536
QY 1133 NGHMV-----YGLQQLNGEVQYFLSNGVQLRESFLENADGSKNYFCHLGNRY 1180
Db 1537 GKQVLSNVKSGTKTDAAVNOQLN-EVRNLLGLG-----NAGNDNADGNQVNIADIKQDN 1591
QY 1181 NGYYS-----FDNDSKWRYFPDASGVMAVGLKTINGNTQYFDQDGY---QVK 1223

Db 1592 SGSSSNRTVIKAGTVLGGKGNNDTEKL-----ATGGIQGV-----DKDGNANGDLS 1638
QY 1224 GAWI--TGSDGKGR-----YFDDGSGNMAVNRFA--DKNGWDYVYVNSDIALVGVQTING 1275
Db 1639 NVWVKTKQDGSKKALLATYNAAGQTNLYLTNNPABAI--INBQIRFFHVNNDQ 1692
QY 1276 KTYVFGQDKQIKGKIITDNGKLYFLA-----NSGELARNIFATDSQNNWYVFGSDGV 1329
Db 1693 EPVVQGRNG-----IDSSASGKHSVAIGFOAKADGEAAVAIGRQTOAGNQSAIGDNA 1745
QY 1330 AVTGSQTI-----AGKLYFPASDKQVKGSPVTVNGKHYHADSGELOVRFEADK 1381
Db 1746 QATGDSQIAITGTVNAGKHSIGAIDPSTVKAD-----NSYSVGNNOFTDATQTDV 1797
QY 1382 DG---NWYVLDNSGEALTGSORINDQRVFFTRGKQVK 1416
Db 1798 FGVGNITVTESNVALGSNAISAG-----THAGTQAK 1831

RESULT 6

US-11-013-759-13

; Sequence 13, Application US/11013759

; Publication No. US20050249747A1

; GENERAL INFORMATION:

; APPLICANT: Loomore, Sheena M.

; APPLICANT: Sasaki, Ken

; APPLICANT: Yang, Yan Ping

; APPLICANT: Klein, Michel H.

; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE

; FILE REFERENCE: PROTEIN OF MORAXELLA

; CURRENT APPLICATION NUMBER: US/11/013,759

; CURRENT FILING DATE: 2004-12-16

; PRIOR APPLICATION NUMBER: US/09/361,619

; PRIOR FILING DATE: 1999-07-27

; NUMBER OF SEQ ID NOS: 32

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 13

; LENGTH: 1992

; TYPE: PRT

; ORGANISM: Moraxella catarrhalis

US-11-013-759-13

Query Match 3.8%; Score 280; DB 7; Length 1992;

Best Local Similarity 19.2%; Pred.No. 5.4e-07;

Matches 307; Conservative 206; Mismatches 603; Indels 482; Gaps 69;

QY 17 VTVAVASGLITLTGTTLGSVSAAETQQTSDKVVVTKQSEDDKA-----ASSEST 66
Db 518 ISVDAKSGNVAPTNYNIGVKTTELNSDGTSDKFSVKGSGTNNLSLVTAHLASVYNEVNR 577
QY 67 DAPKTKQAOTEQTAQASQANVADTSTSIK-----ETPSONITTOANSDDKTVNTK 118
Db 578 -ADSALQSFVKEEDDDANAITVAKDTTKNAGVSIKLGKNGLTVATKKDQTVTFGL 636
QY 119 SEEAQTSSEERTKQSEBAQTASSQALTOAKAELTKQRTAAQENKRPVDLAIPNVKQID 178
Db 637 SQDSGLTIKSTLNNDLTVKDTNEQIQVANGIKFTNV---NGSNPG---TGIANARIT 691
QY 179 -GKYYYIGSDGQPKNFALTNNKVLVFDK-----NTG-----AL 212
Db 692 RDKIFAGSDG-----AVDTNKPYLDDQKLQGVNKKITNTGINAGKAITGSLPLSI 745
QY 213 TQTSQVQFKQGLTKLNNDYTPHNOIVNFENT--SLETIDNYVTADSWYRPKDIILKNGKTW 270
Db 746 ADQSSRIEIGNTIQQDKKNAASINDILNTGFLKNNNNPIDFVSTYDIVD-FANGNAT 804
QY 271 TASSEDLRPLLLMSWPDKQTOIAY-----LNYMNOQGL 304
Db 805 TATVTHDT-----ANKTSKVYDVNVDDTTIHLTGTDNKKLGKVTYKLNKTSANG- 855
QY 305 GTGENYVTADSSQESLNLAOTVQVKIETKISQTOQTQWLRLDI----- 346

Db 1031 KNPTGSSQIQVAGADVGFKAQNNNG-----VVGAGIDGTTTRTRDE-----IGFTGTNGSLD 1082
Qy 452 ANDPEANFDGVRVDAVDNADLLQIASDYLLKAHYG-----VDKSEKNA 495
Db 1083 KSKPHLSKDGINAGGKKTITNQSEIAQNSHDAVTGGKIYDLKTELENKISSSTAKTAQNS 1142
Qy 496 INHLSILEASDN-----DPQNKDTKGAQLPI-----DNKURLSLLYALTRPLEKASKN 547
Db 1143 LHEFSVADEQGNFTVSNPYSSYDTSKTSDDVITPAGENGITTKV-----NKG 1189
Qy 548 EIRSGLE-----PVLITNSLNRSASGKNSERMANVIFIRAHSEVQTVIAKILKAQIN 600
Db 1190 VVRGIDQTKGLTTPKUTVGNNN-----GKGI-----VIDSQNG 1223
Qy 601 PKT-DGLTFTLDELKQAFKIYNEQMRQAKKKTQSNIPATAYALMLSNKSDSTRLLYGDMY 659
Db 1224 QNTITGLSNTL-----ANVNDKGSVRTTEQGNII 1253
Qy 660 SDDQYNMATKSPYYDAIDTLKARIKYAAGQDMKITYVEGDKSHMDWDYTVGLTSVRYG 719
Db 1254 KDEDKTRAA-----SIVDLASGFLQNGEAV-----DFVSTYDVTNFA 1293
Qy 720 TGANEATDQSEATKTKQMAVITSNPSLKNQNDKVIYNMGAHKNQBYRPLLLTTKDG 779
Db 1294 DG-NATT-----AKVTYDDTS-----KTSKVYVDVNVDDTTIEVRDKKLGVK-- 1334
Qy 780 LTSVTSDAASLYRKTD--KGLVFDASDIQGLNPQVSGYLAVVVPVCGASDNQDVRVA 838
Db 1335 TTTLTSTGTGANKFALSNOATGALVKASDIVAHLN-TLSG-----DIQTA 1379
Qy 839 --ASNKANATQVYVESSEALDSOLIEGFSNFQDFVTKDSYTNKKIAQNVQLFKSGVGT 896
Db 1380 KGASQANNSAGYV-----DADGNKVIYD-----STDNKY----- 1408
Qy 897 SPFMAQYVSEDSGFLDSIIQNGYAFEDRYDLAMSKNKYGSQQDMINAVKALHKSIGQ 956
Db 1409 -----YQAKNDGT-----VDKTEKVAKDLV 1429
Qy 957 VIADWYPDQIYNIPGKEVVTATRVNDYGEYRKDSEIKNTLYAANTKSNKGDKYQAKYGGAF 1016
Db 1430 AQAQTPDGTILAQNNVKSVINKEQVNDAN--KKQINEDNAPVKGLEKASADNKTUAAVT 1487
Qy 1017 LSEL--AAKYPSTIPNRTQISNGKKIDPSEKITAWKAKYFNGTNIILRGVGYVLKDNASDK 1074
Db 1488 VGLDNVAQPTLTFAGDTGTAKKL--GETLTI-KGGQDTNKLTDNNIGVW-----AGTD 1540
Qy 1075 YFELKNGQTVLPKQMTNKEA--STGFVNDGNGMTFYSTSYQAQNSFVQDAKGNWYFDW 1132
Db 1541 GFTVK-----LAKDLTNLNSVNAAGTKIDDKGYSFVDSGQAKANTPVLISANG-----LDL 1591
Qy 1133 NGHWV-----YGLQQLNGEVOYFELSNQVQLRESFLENADGSKNYFCHLGNRYS 1180
Db 1592 GKGVISNVGKGTDTDAANVQQLN-EVRNLLGLG-----NAGNDNADGNQVNIADIKKDPN 1646
Qy 1181 NGYYS-----FDNDSKWRYFDASGMVAVGLKTINGNTQYFDQDGY-----QVK 1223
Db 1647 SGSSSNRTVIKAGTVLGGKGNNDTEKL--ATGGIQGV-----DKDGNANGDLS 1693
Qy 1224 GAWI-TGSDGKR-----YFDDSGMNVNRPAN--DKNGDWTYLSNDSGIALVGVQTING 1275
Db 1694 NVWVKTKQDGSKALLATYNAAGOTNYLTNNPAAIDR-----INEQGIIRPFHVDNGQ 1747
Qy 1276 KTYVFGDQKQIKGKIITNGKLKYFLA-----NSGELARNIFATDSQNNWYFSGDGV 1329
Db 1748 EPPVQGRNG-----IDSASGKSHVAIGFOKADGEAAVAIGROTQAGNOSIATGDNA 1800
Qy 1330 AVTGSQTI-----AGKKLYPASDGKQVKGFSVTVYNGKVHYHYHADSGELOVNRPEADK 1381
Db 1801 QATGDQSIAIGTGVNVAGKSHGAIGDPSIVKAD-----NSYSVGNNNQFTDATQTDV 1852
Qy 1382 DG---NWYLDNSNGEALTGQSQRINDORVFTTREGKQVK 1416
Db 1853 FGVGNITVTESNVALGSNSAISAG-----THAGTOAK 1886

RESULT 8

US-11-013-759-7
; Sequence 7, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; FILE REFERENCE: 1038-921MIS:jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 2047
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-7

Query Match 3.8%; Score 280; DB 7; Length 2047;
Best Local Similarity 19.2%; Pred. No. 5.6e-07;
Matches 307; Conservative 206; Mismatches 603; Indels 482; Gaps 69;

Qy 17 VTVAASGLITLTGSSVSAETEQTSDKVVVTKQSEDDKA-----ASESSQT 66
Db 573 ISVDASGNVTAFTYNIQVKTTELNSDGTSDKSVKSGTNNLSLVTAELHSLVINEVNR 632
Qy 67 DAPKTKQAQTEQQAQSAQNAVADTSITK-----ETPSQNTITQAQNSDKTIVNTK 118
Db 633 -ADSALQSEFTVKEDEDDANAITVAKDTTKNAGAVSILKLGKNGLTVA TKDGTVTFFGL 691
Qy 119 SEBAQTSERTKQSEEAQTTASSQALTOAKAELTKQRTAAQENKQPVDLAAIPNVKQID 178
Db 692 SQSGUIGKSTLNNQGLTVKDTNEQIQVGANGIKFTNV---NGSNPG--TGIANARIT 746
Qy 179 -GKYYYIGSDGPKKNFALTNNKVLVYFK-----NTG-----AL 212
Db 747 RDKIGFAGSDG-----AVDTNKPYLDQDKLVGNVKNITNGINAGGKAITGLSPTLPSI 800
Qy 213 TDTSQVQFKQGLTKLNDYTPHNQIVNFENT--SLETIDNYVTADSWYRPKDILKNGKTV 270
Db 801 ADQSSRNIELGNTIQDKKSNAAASINDILNTGPNLKNNNNPIDFVSTYDIDV-PANGNAT 859
Qy 271 TASSESDLRPLLSWWPDKQTQIAY-----LNYMNOOGL 304
Db 860 TATVTHDT-----ANKTSKVYVDVNVDDTTIHLTGTDNDKKLGKVTTKLTKTSANG- 910
Qy 305 GTGENTADSSQESLMLAAQTQVQKLETKISQTOQTQWLRLDI----- 346
Db 911 NTATFNVNSDEDLAVNAKDAENLNTLAKEHTTKGTADTALQTFVTKVVDENNADD 970
Qy 347 INSFVKTPNWSQTESDTSAGE-----KDHLOGGALLYSNSDKTAYAN---SDYRLLN 397
Db 971 ANAIVTQXKANNQVNTLTKGNGENLNIKTDKNGTVTFGINTTSGLKAGKSTLNDGGLSI 1030
Qy 398 RPTPS-----QTKGP--KYFEDNSSGGYDFELLANDINSNPNVQAEQLNHLHLYMVGSI 451
Db 1031 KNPTGSEQIQVAGADVGFKAQNNNG-----VVGAGIDGTTTRTRDE-----IGFTGTNGSLD 1082
Qy 452 ANDPEANFDGVRVDAVDNVDNADLLQIASDYLLKAHYG-----VDKSEKNA 495
Db 1083 KSKPHLSKDGINAGGKKTITNQSEIAQNSHDAVTGGKIYDLKTELENKISSSTAKTAQNS 1142
Qy 496 INHLSILEASDN-----DPQNKDTKGAQLPI-----DNKURLSLLYALTRPLEKASKN 547
Db 1143 LHEFSVADEQGNFTVSNPYSSYDTSKTSDDVITPAGENGITTKV-----NKG 1189

609 Db -----HYGSGTGTSMASPOIAGASLLVQYLEKTQPNLPKEKADIIVKNILMSNAQIHVN 663
767 QY -----LLLTTKOGLTS-----YTSAAAKSLYRKT 796
664 Db PEKTTTSPROGAGLLNIDGAVTSGLVYVTKDNYSISIGNITDITDFTVTHVNLNK- 722
797 QY NDGELVFDASDIQYINLPQVGYLA VWPVGASDNQDVRVAASNK--ANATQGVYESS 854
723 Db -DK-TLRYDTLLELTDHVDPOKGRFTLTSLKTYQGGVTPANGKVTVRVTMDVSFTK 780
855 QY ALDSQL-----IYBGSFNQD-----FVTKDSYTNKKIA-QNVLPKSGVTS 897
781 Db ELTKQMSNGYLLGFRVFRDSQDLNRVNPFFVGKQGFENLAVAEESYRLKSGQKTG 840
898 QY F--ENAPQ---YV-----SSEDSFLL----- 913
841 Db FYDESGPKDDIYVGRHFTGLVLTGSETNVSTKTIISDNGLHTLTGPKNADCKFILEKNAQ 900
914 QY -----DPIIQNGYAFEDRYDLAMSKN---NKY-GSQQDMINAVKALHKSQIQTADWVPDQ 965
901 Db GNPVLAISPNG---DNNQDFAAFKGVFLRYQGLKASVYHASDKEHKNPL-----WVSP 952
966 QY IYNLPKEVVTATRVNDYGEYRKDSEIKNTLYAANTKSNKDYQAK-YGGAFLSE----- 1019
953 Db SFX-----GDKNFNSDIR---FAKSTTLGTAFSGKSLTGAELPDGYHYH 994
1020 QY LAAKYPSPF--NRTOISNGKKIDPSEKITA WAKAKYFNGTN-----ILRGVGVYVLKDNA 1071
995 Db VVSYYDDVGVAKQEWTFMILBRQPVLS-QATFDPETNRFPPELKDRLGAGVRKDSV 1053
1072 QY SDKYFELKNGQTYLPKQMTNKEASTGFVNDG-----NGMTFYSTSGYQAKNSFV--Q 1121
1054 Db F--YLERKONKPY-----TVINDSYKYVSVEDNKT FVER---QADGSFILPLD 1097
1122 QY DAK-GHWY--FDNN-----GHMVTGLQOLNGEVQY----- 1149
1098 Db KAKLGFYFYMVEDFAGNVAIAKLGDHLPQTLGKTPIKLITDGNVQTKETLKNLEMTQS 1157
1150 QY ---FLSNGVQL-----RESFLE-NADGSKNYFGLH--NRYSN---GY 1183
1158 Db DGLVNTQOALVVRHNPQSQSLTKVKNQDFFISPNEDGNKDFAFKGLKNVYNDLTVNV 1217
1184 QY YSFNDNSKRYFDAS-----GVNAVGLKTINGNTQYF---DQDGYQVK 1224
1218 Db YAKDDHQKTPTWSSQAGASASAIESTAWGITARGSKVMPGQYQVTVYRDEHGEHQ 1277
1225 QY AMITGSDGKRYPDDGSGNM--AVNRPANDKNGDWYLLNSDGIALGVQVOTINGKTYFGQ 1282
1278 Db QYTVISVNDKKPMITQGRFDTINGVDHFTPKT---KALGSSGI-----VREBEVFLAK 1327
1283 QY -----DGKQIKGIITDNGKLVFLANSDELARNIFATD--SONNWWY----- 1323
1328 Db KNGRKFVDTEGKD--GITVSDN---KXVTPKXPDGYSYITSKRGVTLSDYYLVEDRAGN 1382
1324 QY -----FGSDGVAVTGSGQTIACKKLYFASDGKQVKG----- 1353
1383 Db VSFATJRLDKAVGKDKAVNVFGLDLPVPEDKQIVNFTYLRVADGKPIENLEYNNSGNS 1442
1354 QY -----SFTYNGKHYHYHADSGELQVNR---FEADKQNGWYLLNSGNGBALTSQ 1399
1443 Db LILPYGKTYVELLTYDTNAKLESKDIVSFLLSADNNFQVTPKMTMLATSQ 1494

RESULT 10

US-11-013-759-11
; Sequence 11, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.

; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; FILE OF INVENTION: PROTEIN OF MORAXELLA
; FILE REFERENCE: 1038-921MIS:jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 2314
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
; US-11-013-759-11

Query Match 3.7%; Score 275; DB 7; Length 2314;
Best Local Similarity 20.3%; Pred. No. 1.2e-06;
Matches 307; Conservative 192; Mismatches 598; Indels 418; Gaps 66;
QY 13 KKHVTVAVASGLITLTGTTTLGSSVSAETEQQTSQDKVVTQKSDEDDKAASESSQTDAPKTK 72
Db 367 KKLGETLTIKGGETQADKLTDDNNIGVVDNNTGLKV-----KLAKNLGLETVSTK 418
QY 73 Q-AQTEQTAQQAQNVADTSTSIKETPSQNIITQANSDDKTVTNTK-----SBEAQSTEE 127
Db 419 NLTASEKVTVGSGNNTAELQSGGLTFTPTNAST-----DKTVYGTDLKFTDNTALE 473
QY 128 RYKQSEEAQTASSQALT--QAKAELTKQRTAAQENKPVDLAALPNVKQIDGKYYIG 185
Db 474 DTTRITKDKIGFSNKGAGTVDENKPYLDKDKLVGNSTLANGGL-TVNTTIGGSNKKQIQVG 532
QY 186 SDQPKKPFALTVNN-----KVLVFDK--NTGALTDTSQ 217
Db 533 ADGIRFADVNVNVAAKFGTTRITEEIGFADADGKVDKPSYLDKKQLQVGVKITKD 592
QY 218 YQFKQGLTKLNN-----DYTPHNQIVNFENTSLETINDVYVAD--SWYRPKDILKN 266
Db 593 SGINAGDQKISNVKDATDDTDAVYKQLQVQDADGALQSPSIRDEKQGQEFISNLYSN 652
QY 267 GKTWTASSEDLRPLLMWHPKQTOIAYLVNMQOGLGTGENYTDSSQESLNLAAQTV 326
Db 653 GNT-----PNTFETITF-----AGENGISIN--DIARGKV 681
QY 327 QV-----KIETKISQTOQWLRDIINSFVKTOPNMNSQTESDT 365
Db 682 KVGIDPGLNLTTPKLTGVSQDKGQTLVIEQVAGSNDTKNIIIRGLSPTLP-----SIT 734
QY 366 SAGEKDHQGLGALLYSNDSKTAYA-----NSDYRLNLRPT-----SQ 403
Db 735 NAGGVRTTEQNTITSDEDESKAASIGDILNTGFNLKNNNSVGFVSTYNTVDFIDGNAT 794
QY 404 TCKPKYFEDN--SSGGYD-----FLANDIDNSNPV-VQAEQLNLWHLHYLMYGSIVAN 453
Db 795 TAKVYDETNTQTSKVTYDVNVNDEKTIETGDNKGNKIKGIVKTTTLTTN-----AN 845
QY 454 DPEANFDGVRDVAQVNVNA-DLLQIASDYLKAHYGVDKSEKNAIHLNLSILEAWSNDPOY 512
Db 846 GKATNFTSTNDAL--VNAKDIAENLNTLAKIHTTKGTADTALQTFKVKKGCGATDETI 903
QY 513 NKDTKGAQI-PIDNKLRLSLLYALTRPLEKDSASNKNEI--RSGLEPVIITNSLNRSAGK 569
Db 904 TVGDKGTQNGKTVNTLKLKGENGLTVATNKGDTVTFGINTQSGLKAGDSTTLNKDGLSIK 963
QY 570 NSERMANYIFIRAHSEVQ-----TVIAKIIKAQINPKTDGLT----- 607
Db 964 NP-----ASNEQIQVGADGVKFAKVDKXNSSTGIDGTSRITKQDQIGFTGANGSLD 1013
QY 608 -----FTLDELKQA--FKYINEDMRQAKKYQTSQNTPTAVALMLSNKDSIT----- 651
Db 1014 TTKPHLTQDKLVGEVEITNTGINAGGKAIT--NQSGDITQNSN-DATVGRVYDLKTE 1070
QY 652 -----RLYYGDMYSDDQVYMATKSPY--YDAIDTLLKARIKYAAGQDMKI 695

Db 1071 LESKINSAKTAQNSLHPSVADEQGNHFTVSNPNYSYDTSKT--SDVITFA--GENGIT 1126
Qy 696 TYVEGDKSHMDWYTGVLTSVRYGTGANBATDQGSBATKTQGMVITSNPNLSKLQNQDK 755
Db 1127 TKVNGVWRVIGIDQTKGLTPKLTUVGNNG--KGIVDSKQDQNTITGLSNTLANVTND- 1183
Qy 756 VIVNMGAAHKNQYRPLLLTTKDGLTSTYSDAAKSLYKRTNDKGLVFDASDIQGYLNP 815
Db 1184 -----GAGH-----ALSQGLANDTDKTRAASI-----GDTLANAGFNLOQ--NG 1219
Qy 816 QVSGYLAVVPVVGASDNQDVRVAASNKANA--TCQVYESSALDSQLIYEGFSNFQDFVTK 874
Db 1220 EAVDFVSTYDVDFID-----GNATTAKVYDTSKTSKVY----- 1256
Qy 875 DSDYTNKKIAQNVLKFSKMGVTSFEMAPQVYSSSDGSLDSIIQNGYAFEDRYDLAMSKN 934
Db 1257 DVNVNDKTI--EVTSDKGLGVKTTLTK--TSANGNATKFSAADGDALVKASDIATHLN 1311
Qy 935 NKYGSOODMINAVKA-----LHKSGIOVIADWVPQIYNLPKGEVVTATRVNDYGEYRK 988
Db 1312 TLAGDIOQAKGASQASASSAYVDADGNKVIYDSTDKKY-----QVNDKGQVDK 1360
Qy 989 DSEI-KNTLYAANTKNGKDYQAKYGAFLSELAAYPSIFNRTOISNGKKIDPSEKITA 1047
Db 1361 NKEVAKOLVA-----QAQTPDGTIAQNVK--SVINKEQVND-----A 1397
Qy 1048 WKAKYFNGTNILGRGVGYVYLKDNASDKYPFELKGNQTYLPKQMTNKEASTGFVNDGNGMTF 1107
Db 1398 NKKQGINEDNAFTIKGLENAAKDTKKNAAVTVGDLNAVAQTPLTFAGDTGTTAKKLGETL 1457
Qy 1108 YSTSGVQAKNSFVQDAGKNWYFDDNNGHMYGLQQLNGEVOYPLSN-----GVOLRE- 1159
Db 1458 TIKGGQTDNKL-T-----DNNIGVVAGTDFGFTVLAKDLTNLNSVNAAGGTIDEK 1507
Qy 1160 --SFLNADG-----SKNYFCHLGNRYSN-GYVSPDNDG-KWRYFDASGMAVGLK 1206
Db 1508 GISFVD-ANGQAKANTPVLISANGLDGGKRIISNIGAAVDNDNAVNFQKNEVA-----K 1560
Qy 1207 TINGNTQYFDQGYQV-----KGAWITGSDGKKRYFDDGSGNMVNRFPANDKNGDWY 1259
Db 1561 TVN-NNLNSQNSGASLPFVVVDANGKPIGTDGKPKQKAIGA-----DGKYYH 1607
Qy 1260 LNSDGLALVQVQTINGKTYFGDQKQIKKIITDNGKLYFLANSSEL--ARNIATDS 1317
Db 1608 ANANGVPV-----DKDGP-----ITDADKLANLAAGKPLDAGHQVVASLG 1649
Qy 1318 QNNWYFGSDGVAVT 1332
Db 1650 GN-----SDAITLT 1658

RESULT 11
US-11-052-554A-231
; Sequence 231, Application US/11052554A
; Publication No. US2005028886A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 231
; LENGTH: 701
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae R6
US-11-052-554A-231

Query Match 3.6%; Score 270; DB 7; Length 701;
Best Local Similarity 19.9%; Pred. No. 4.4e-07;
Matches 178; Conservative 114; Mismatches 285; Indels 318; Gaps 37;
Qy 470 VNADILLQIASDYLKAHYGVDSKNAIHLSTLEAWSNDPOYNKDTKGAQLPIDNKLRL 529
Db 89 LNIKLSAITSKTYLR-----ELNVLE-----EKSD-----ELPSEIKAKL 123
Qy 530 SLLYA-----LTPLEKDAASKNEIRSGLEPVITNSLNRSAEGKNSEMANIYIFIRAH 584
Db 124 DAAFEKFKDULKPGKVAEAKKVEEA-----KKKAEDQKEEDRRNY-----PT 168
Qy 585 SEVQIVIAKIIKAOINPKTDGLTFLDELKQAFKIYNE--DMROAKKKTQSNIPATAYALM 643
Db 169 NTYKTELEIEAFDFVKVKEAELELVKEBEAKS---RNEGTTIKQAKEK-----V 213
Qy 644 LSNKDSITRLYYGDMVSDDGQVMATKSPYDAIDTLLKARIKYAAGGQDMKITVYVBGDKS 703
Db 214 ESKKABATRL-----ENIKTDKKAEBEAKKADAKL----- 245
Qy 704 HMDWYTGVLTSVRYGTGANEAAT--DQGSBATKTQGMVITSNPNLSKLNDKVINMGMA 762
Db 246 -----KEANVATSDQ--KPKGRA----- 262
Qy 763 AHKNOEYRPLLLTTKDGLTSTYSDAAKSLYKRTNDKGLVFDASDIQGYLNPQVSGYLA 822
Db 263 ----KRGVPELATPD-----KKEND----- 279
Qy 823 VWPVPGASDNQDVRVAASNKANATQVYESSALDSQLIYEGFSNFPDQFVTKDSYTNKK 882
Db 280 -----AKSDSSVGEEETLPSLSLKGKGVAAEAKKVEAEKAKAD----- 319
Qy 883 IAQNVLKFSKMGVTSFEMAPQVYSSSDGSLDSIIQNGYAFEDRYDLAMSKNKGYSQDD 942
Db 320 ---QKEEDRRNYPTNTYKTLDEIAESDVKVKEAELE-----LVKEBAKEPRDEE 366
Qy 943 MINAVKALHKSQIQUIADWVPQIYNLPKGEVVTATRVNDYGEYKDSIKNTLYAANTK 1002
Db 367 KIKQAKAKVES-----KKAEATRLENIKTDKKAEBE-----AKRK 402
Qy 1003 SNGKDYQAKYGAFLSELAAYPSIFNRTOISN--GKKIDPSEKITAWKAKYFNGTNILG 1060
Db 403 AAEBEK-----VKEKPAEQPAPATQPKPAPKPEKPAEQPKAEKT----- 444
Qy 1061 RGVGYVLKDNASDKYFELKGNQTY--LPKQM--TNKEA-----STGFVNDGNGMTFYST 1110
Db 445 -----DDQAEEDYARRSEBEYNRLTQQPPKTEKPAQSPKTPKTMGKQENGMMWYFNT 497
Qy 1111 SGYQAKNSFVQDAGKNWYFDDNNGHMYGLQQLNGEVOYFSLNGVQLRSFLENADGSKN 1170
Db 498 DGSMA--TGWLQ--NGSWYILNANGAMATGWLQNGSWYIL-----NANGS-- 540
Qy 1171 YFGHLGNRYSGYVSPDNDGSKWRYFDASGVMAVGLKTINGNTQYFPDDG-----YQVK 1224
Db 541 -----MATGW--LQNGSWYILNANGAMATGWLQNGSWYILNANGAMATGWLQYNG 590
Qy 1225 AWITGSDGKKRYFDDGSGNMVNRFPANDKNGDWYILNSDGIALVGQVQTINGKTYFGDQ 1284
Db 591 SW-----YYLANGDMATGWLQN--NGSWYILNANGDMATGWLQNGSWYILNANG 639
Qy 1285 KQIKGIITDNGKLYFLANSSELARNIFATDSQNNWYFSGDGVAVTCSQITAG 1339
Db 640 DMATGWV--KDGDTWYILEASGAMKASQWFKVS--DKWYVNGSG--ALAVNTTVDG 690

RESULT 12
US-11-022-562-228
; Sequence 228, Application US/11022562
; Publication No. US20050249742A1
; GENERAL INFORMATION:
; APPLICANT: Ruprecht, Ruth M.
; APPLICANT: Shibeong, Jiang

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING
; FILE REFERENCE: A CYTOTOXIC T LYMPHOCYTE IMMUNE RESPONSE
; CURRENT APPLICATION NUMBER: US/11/022,562
; PRIOR FILING DATE: 2004-12-22
; PRIOR APPLICATION NUMBER: PCT/US03/20322
; PRIOR FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: 60/392718
; PRIOR FILING DATE: 2002-06-27
; NUMBER OF SEQ ID NOS: 340
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 228
; LENGTH: 396
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-11-022-562-228

Query Match 3.6%; Score 267.5; DB 7; Length 396;
Best Local Similarity 26.7%; Pred. No. 2.8e-07;
Matches 94; Conservative 45; Mismatches 124; Indels 89; Gaps 16;
QY 1051 KVFNGNIIIRG-VGVVLKDNASDKYFELKGNQYLPKQMTNKEASTGFVNDGNGMTFFYS 1109
DB 91 KSFENSELDRDLHGLKIDNKTYYYDEA-----SKLVGLININNSLFYFD 137
QY 1110 -----TSGYQAKNSFVQDAKGNWYFDD-NNGHMVYGLQOLNGEVOYFLSNGVQLRESEL 1162
DB 138 PIESNLVTGQNTN-----GKKYFDDINTGAASTSYKIINGKHFFYFNNGV-MQLGVF 189
QY 1163 ENADGSKNFGHLGNRYN-----GYYSFDNDSKWRFYDASGVNAVGL 1205
DB 190 KGPDPGE-YFAPANTQNNIEGQAIYVQSKFLTLNGKTYFDDNSK-----AVTGW 239
QY 1206 KTINGNTQYFDQDGVQVKGAWITGSGKKRYFDD-----GSGNMAVNRFANDKNGDWYYL 1260
DB 240 QIIDGKYYFNLTAERATGWQT-IDGKKYFNTNTSIASGTII-----NGKHFFY 291
QY 1261 NSDGIALVGV-QTINGKTY--FGQDGKQIKGKIITDNGKLYFLANSGLARNIPATDS 1317
DB 292 NTDGIWQIGVFGPGNGFEYFAPANTDANNIEGQAIYQNRFLY----- 335
QY 1318 QNNWYFSGDGVAVTGSQIAGKLYFASD-GKQVKGSVFTYNGKVHYHAD 1368
DB 336 HDNIYFNGNSKAVTGWQTINGNNGVYFMPDTAMAAAGGLFEIDGVYIFFGVD 387

RESULT 13
US-10-793-626-2964
; Sequence 2964, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2964
; LENGTH: 5024
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (5024)
; OTHER INFORMATION: variable amino acid
US-10-793-626-2964

Query Match 3.6%; Score 266; DB 6; Length 5024;
Best Local Similarity 18.9%; Pred. No. 1e-05;
Matches 259; Conservative 211; Mismatches 561; Indels 340; Gaps 59;
QY 18 TVAVASGLITLGTITLGSSVSARETEQTSQVKTQ-----KSEDDKAAESSQSDA 68
DB 1060 TINQLANAVTQAKSNLHGDTKLQHDKDSAKQTIAQLQNLNSAQKHMEDSLIDNESTRT-- 1117
QY 69 PKTKQAQTEQTOAQS-----QANVADTSTSTITKTPS-----QN 102
DB 1118 -----QVQHDLTEAQALDGLMGALKESIKONTNIVSNGNYINAE-PSKKQYDAVQAQNAQN 1172
QY 103 ITTQAN-----SDDKTVTNTKS-----BEAQTSERTKQS-----BEAQTTAS 140
DB 1173 IINGTNOPTINKGNVTTATQTKNTKDALDGDHRLSEAKNNANQTIIRNLNLNAQKDAE 1232
QY 141 SQALTOAKA-ELTKQRTAAQENKPNVD--LAAIPNVKIDGKYIYIGSGQPKNF--- 194
DB 1233 KNLVNSASTLEQVQQLQTAQQLDNAMGELROSIAKNDQVKADSKYLNEDPOIKQNYDDA 1292
QY 195 -----ALTVNKVLVYFDKNTGALTDTSQYQFKOGLTKLNNDYTPHQNIVNFPNTSLE 246
DB 1293 VORVETIINETQNPPELLKANIIOATOSVQNAEQALHGAELKQD-----KOTSST 1342
QY 247 TIDNYVTADSWYRPDKILKNGKTTWASSESRLRPLMSWMPDKQTQIAYLN----- 297
DB 1343 ELDGLTDLTDAQREK---LREQINTNSRDDIK-----QKIEQAKALNDAMKKLKEQ 1391
QY 298 YNNQGLGTGENYNTADSS--QBSNLAAOTVQVKIETKISQTOOTQWLRIIDINSFVKTP 355
DB 1392 VAQKDCGVHANSYDYNEDSAQDAYNNALQAEIINNNSPNLNAQDIITNALNNIKQAQD 1451
QY 356 NNNS-----QTESDT---SAGEKDH---OGGALLYNSDKTA----- 387
DB 1452 NLHGAQKLQDDKNTTNOAIGNLNLNQPOKDALIOAINGATSRDQVAELKEAEALDEAM 1511
QY 388 -----YANDYRLLNARTP--TSQTGPKYFENDSSGGYDFLLANDIDN--SNPVQAEQL 438
DB 1512 KQLEDQVNDQDQISNSSPINEDSDQKTYNDKIQ-----AKEIINTSNTLQDKKI 1565
QY 439 -----NWLH-----YLMYGSIVANDPEANFGV--RVDADVNVNA 472
DB 1566 ADTLQNIKDAVNNLHGDQKLAQS KODANNQNLHLDLTBEQKNHFKPLINNADTRDEVNK 1625
QY 473 DL---LQIASDYLKAHYGVYDKSEKNAIHLNLS--ILEAWSNDNDPOYKNDTKGAQLPIDNKL 528
DB 1626 QLEIAKQLNGDMSTLHKVI--NDKQIOIHLNSYINADNDKQNYDNAIKEADELIHHPD 1683
QY 529 LSLLYALTRPLEKADASNKEI--RSGLEPVTINSLNRSAGKNSERMANYIFIR---AH 583
DB 1684 TLDHKALQDLINKIDQAHNELNGESRFKQALDNALND--IDSLSLNVPQRTVKDNINH 1741
QY 584 DSEVOTVIAKIIKAQINPKTDGLTFTLDELKQAFK-----IYNEDMRQAKKYQTQSNII-- 636
DB 1742 VTTLESQAQELQAK-----ELNDAMKWRDSIMNQEIQIRKNSNYTNEDLAQ 1788
QY 637 PTAYALMLNSKDSITRELYYGDYMSDDQYMATKSPYDAIDTLLKARIKYAGGQDMKIT 696
DB 1789 QNAYNHAVDNIINNII-----GEDNATMDPQI-----IKOATQDINTATN 1827
QY 697 YVEGDKSHMDMDYTGVLTSVRYGTGANEATDQ-----SEATKTOGMVITSNPSSL--K 749
DB 1828 GLNGDKQLQD-AKTDAKQIITNFTGLTPEKQALENIINQOTSRANVAKQLSHAKFLNGK 1886
QY 750 LNQNDKVIYNMGAHKNQYRPLLLTTKDGLSYTSDAAKSLYRKTNDKGLVDFDASDI 809
DB 1887 MEELKVAVAKASLVRQNSY-----INEDVSEKEAYEQAIKAGGEIINSEN- 1932
QY 810 QGYILNPQVSGYLAVWVPVGSQNDQVRVAASKANATQGVYESSALDSQLI-----Y 862
DB 1933 -----NPTIS-----STDINRTIQEINDAEQNLHGENKLRAQAEIAKNEIQNL 1975
QY 863 EGFNSFQDFVTK---DSDYTNKKIAQNQLPKSWGVT-SFEMAPQVVSSEDSGFLDSIIQ 918

Db 1976 DGLNSAQ--ITKLIQIDGRTTKPAVTKLEBAKAINQAOQLKQSIADKDATLNS-- 2030
QY 919 NGYAFEDRYDLAMSKNNKYGSDQDMINAVK--ALHKSQIGVIADWVPDQIYNLPKGEVVT 976
Db 2031 -NYLNEDS-EKKLAYDNVSAEQELINQLNDPTMDISNIOAITQKVIQAKDSLHG----- 2083
QY 977 ATRVNDYGEYRKDSEIKNTLYAANTKSNKDVQA-----KYGGAFLSELAAKYPSIFN 1029
Db 2084 ---ANKLAQOADS---NLIIINQSTNLNDKQKQALNDLINHAQTQKQVAEIIAQANKLAN 2137
QY 1030 RTQISNGKKIDSEKITAOKAKYFN-----GTNILGRGVGYVLKON-ASD 1073
Db 2138 --EMGTLTKLVESQSNVHOQSKIINEDPQVQNIYNDISIQKGREILNTGTTDDVLNNKNIAD 2195
QY 1074 KYFELKGNQTYLPKQMTNKEASTGFVNDGNGMTFFYSTGYQAKNSFVQDAGNWWYFDNN 1133
Db 2196 AIQNIHLTKNDLHGDKLQKQAQDATNELNYLTNLNNSQKQSEHDEINSAPSRTEVSNDL 2255
QY 1134 GH---MVYGLQQLNGEVQYFLSNGVQLRESFLENADGSKNYFGHLGNRYSN 1181
Db 2256 NHAKALNEAMROLENEVA--LENSVRKLSDFINEDEAAQ-----NEYSN 2297

RESULT 14
US-11-052-554A-229
; Sequence 229, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT FILING DATE: 2005-02-07
; PRIOR FILING DATE: 2004-07-20
; PRIOR FILING DATE: 2004-07-20
; PRIOR FILING DATE: 2004-07-20
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 229
; LENGTH: 619
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae R6
US-11-052-554A-229

Query Match 3.6%; Score 265; DB 7; Length 619;
Best Local Similarity 19.8%; Pred. No. 6.8e-07; Indels 244; Gaps 31;
Matches 158; Conservative 106; Mismatches 300; Indels 244; Gaps 31;

QY 642 LMLSNKDSITRLYGDYSDGYMATKSPYYDAITLKLARIKYAAGGQDMKITVVEGD 701
Db 6 MILTSLASVAILGAGFVASOPTVRAEESP-----VASQSKAEKDYDAAKKDAK---NAK 57
QY 702 KSHMDWDTYGLTVSRYGTGANRATQGSSEATKTCQMAVITSNPSLKLNDKVINMG 761
Db 58 KAVED-----AQALDDAKAAQ-----KYDEDQKTEEXA 88
QY 762 AAHK--NOEYRPLLLTTKDLGTSY--TSDAAAKSLYRKNTDKG-----ELVPDASDIQY 812
Db 89 ALEKASEEMDKAAVAAYQOATDIAKADKADKIMIDEAKKREBAKTKFNTVRAM 148
QY 813 LNPQVSGYLAVVPVGASNDQDVRVAASNKA-NATQGVYESSALDSQLIYEGFSNPDQF 871
Db 149 VVPE-----PEQLAETKKSEAKQKAPETTKLBEAKALEE----- 186
QY 872 VTQSDYTNKIAQNQLFKSWGVTSFEMAPQVSSE-----DGSFLDSI 916
Db 187 -----AEKKATEAKQ-----KVDABEVAPOAKIAELENNQVHRLBEQLKEIDSESEDY 234
QY 917 IONGY--AFEDRYDLAMSKNNKYGSDQDMINAVKALHKSQIQVIADWVPDQIYNLPKGEV 974

Db 235 AKEGFRAPIQSLDAAKAKLSKLELSDKIDELDA-----EIAK-LEDQL-----KAA 281
QY 975 VTATRVNDYGE-----YRKDSEIKNTLYAANTKSNKDVQYQAKYGGAFLESLAAKYPSI 1027
Db 282 ENNNVEDYFKEGLEKTTIAAKKAELEKTEADLKCAVNEPEKPAPEPETPAPEAPAEQPKP 341
QY 1028 FNRTQISNGKKID-PSEKITAWKAKYFNGTNIILGRGVGYVLKDNASDKYFELKGNQTY-- 1084
Db 342 APAPQAPAPKPEKPAEQPKPEKT-----DDQQAEEYARRSEBEYNR 384
QY 1085 LPKQMTNK-----EASTGFVNDGNGMTFFYSTGYQAKNSFVQDAGNWWYFDNNHMYV 1138
Db 385 LTQQQPPKPAEKPAPEKPTGWKQENGWYFYNTDGSNA-TGWLQN-NGSWYILNSNGAMAT 442
QY 1139 GLQQLNGEVQYFLSNGVQLRESFLENADGSKNYFGHLGNRYSNYYSPDNDKSWRYFDAS 1198
Db 443 GWLQYNG-----SWYILNAN 457
QY 1199 GYMAVGLKTINGNTQYFDQDGYQVKGAWITGSDGKKRYFDDGSGNMAVNRFANDKNGDWY 1258
Db 458 GAMATGWAKVNGSWYILNANGAMATG-WL-----QYNGSWY 492
QY 1259 YLNSDGIALLVGQTINGKTYFYFGDQKQIKGKIITDNGKLKYFLANSSELARNIFATDSQ 1318
Db 493 YLNSANGAMATGWAKVNGSWYILNANGAMA-----TGMLOY-----N 528
QY 1319 NNWYFPGSDGVAVTGTSQTIAGKLLYPASDGKQVKSFTYTYNGKVHYHYHADSGELQVNRPE 1378
Db 529 GSWYILNANGAMATGWAKVNGSWYILNANGAMATG--WYKDGDTWYLYLEASGAMKASOWP 586
QY 1379 ADKDGWYLLDSNGEALTGSQRINDRV 1406
Db 587 KVSD--KWIYVNGLG-ALAVNTTVDGYKV 612

RESULT 15
US-10-873-528-17
; Sequence 17, Application US/10873528
; Publication No. US20050276814A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hanabro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129WO
; CURRENT APPLICATION NUMBER: US/10/873,528
; PRIOR FILING DATE: 2004-06-23
; PRIOR APPLICATION NUMBER: US/09/769,787
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 658
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-873-528-17

Query Match 3.6%; Score 265; DB 6; Length 658;
Best Local Similarity 24.1%; Pred. No. 7.4e-07; Indels 106; Gaps 21;
Matches 106; Conservative 59; Mismatches 168; Indels 106; Gaps 21;

QY 1048 WKAKYF---NGTNILGRGVGYVLKONASDKYFELKGNQTYLPKQMTNKEASTGFVNDGNG 1104
Db 28 WQCKQYLKEDGSGQAANEV---FDTHYSWFYIKADANY-----AENELKQGD 74
QY 1105 MTFYSTGYQAKNSFVQDAGNWWYFDNNHGM-----VTGLQ- 1141
Db 75 YFVLKGGYMAKSEWED-KGAFYYLDQDGKMKRNAWVGTSYVGATGAKVIEDWYDSQY 133

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 19:15:50 ; Search time 29.5158 Seconds
(without alignments)
4661.567 Million cell updates/sec

Title: US-10-797-821-36

Perfect score: 7462

Sequence: 1 METKRYKMKHKHWTVA.....EGKQKGVAYDERRLLVYR 1430

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 80:**

1: pir1:**

2: pir2:**

3: pir3:**

4: pir4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	7443	99.7	1431	2	A45866
2	3817.5	51.2	1475	2	B31335
3	3788.5	50.8	1577	2	T30858
4	3713.5	49.8	1449	2	T30857
5	3713.5	49.8	1449	2	T30552
6	3591.5	48.1	1375	2	JT0345
7	3542	47.5	1592	2	A38175
8	3502.5	46.9	1518	2	A44811
9	3312	44.4	1599	2	S22737
10	3252.5	43.6	1508	2	T31098
11	3244.5	43.5	1290	2	JCS473
12	3174	42.5	1365	2	A41483
13	562	7.5	2817	2	B97033
14	451	6.0	2710	2	A37052
15	435	5.8	563	2	A37184
16	412	5.5	2178	2	S55805
17	393	5.3	2367	2	S70172
18	388	5.2	2364	2	I40884
19	387	5.2	2366	2	S10317
20	363.5	4.9	648	2	S10869
21	306.5	4.1	1463	2	T30290
22	296.5	4.0	690	2	F98114
23	293.5	3.9	2334	2	S32920
24	291	3.9	1335	2	T30211
25	278	3.7	721	2	C97980
26	275.5	3.7	6713	2	B89921
27	272	3.6	1819	2	D97033
28	272	3.6	1881	2	H95076
29	270	3.6	701	2	H98120.

30	268.5	3.6	1946	2	A61449	hypothetical prote
31	265	3.6	619	2	A97887	surface protein ps
32	265	3.6	619	2	A41971	surface protein ps
33	265	3.6	627	2	G97975	hypothetical prote
34	265	3.6	658	2	E95111	endo-beta-N-acetyl
35	263	3.5	2167	2	Af1489	cell wall-associat
36	262	3.5	601	2	S57962	cspC protein - Clo
37	262	3.5	3890	2	C89921	hypothetical prote
38	261	3.5	1385	2	D89824	hypothetical prote
39	260.5	3.5	2285	2	T12796	probable transglyc
40	259	3.5	1296	1	HMS01F	aggregation protei
41	257.5	3.5	340	2	G95043	choline binding pr
42	257.5	3.5	627	2	E95107	hypothetical prote
43	257.5	3.5	2271	2	F90073	hypothetical prote
44	255.5	3.4	1423	2	A49206	exo-beta-D-fructos
45	255.5	3.4	1959	2	AG1085	hypothetical prote

ALIGNMENTS

RESULT 1

A45866

dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans

C:Species: Streptococcus mutans

C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Mar-2004

C:Accession: A45866

R/Honda, O.; Kato, C.; Kuramitsu, H.K.

J. Gen. Microbiol. 136, 2099-2105, 1990

A:Title: Nucleotide sequence of the Streptococcus mutans gtfd gene encoding the glucosyl

A:Reference number: A45866; MUID:91100958; PMID:2148600

A:Accession: A45866

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1431 <HON>

A:Cross-references: UNIPARC:UPI000017AC5C; GB:M29296

C:Keywords: glycosyltransferase; hexosyltransferase

F:181-201/Domain: cpl repeat homology <CP1>

F:1127-1146/Domain: cpl repeat homology <CP2>

F:1192-1211/Domain: cpl repeat homology <CP3>

F:1257-1276/Domain: cpl repeat homology <CP4>

F:1277-1297/Domain: cpl repeat homology <CP5>

F:1321-1340/Domain: cpl repeat homology <CP6>

F:1341-1361/Domain: cpl repeat homology <CP6>

F:1385-1404/Domain: cpl repeat homology <CP7>

Query Match 99.7%; Score 7443; DB 2; Length 1431;

Best Local Similarity 99.9%; Pred. No. 7e-318;

Matches 1427; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy	1	METKRYKMKHKHWTVA	ASGLITLGTTLGSSVSAETQOTS	DKVVTQKSEDDKA	60
Db	1	METKRYKMKHKHWTVA	ASGLITLGTTLGSSVSAETQOTS	DKVVTQKSEDDKA	60
Qy	61	SSSQTDAPKTKQAQT	EQTAQSQANVADTSITKETPSQ	NIITQANSDDKTVNTKSE	120
Db	61	SSSQTDAPKTKQAQT	EQTAQSQANVADTSITKETPSQ	NIITQANSDDKTVNTKSE	120
Qy	121	EAQTSEERTKQSEEAQ	TASSQALQAELTKQRQTAAQEN	KNPVDLAAIPNVKQIDGK	180
Db	121	EAQTSEERTKQSEEAQ	TASSQALQAELTKQRQTAAQEN	KNPVDLAAIPNVKQIDGK	180
Qy	181	YYIGSDGPKNFALT	VNNKVLVFDKNTGALTDTSOYQ	FKGLTKLNNDYTPHNOVNF	240
Db	181	YYIGSDGPKNFALT	VNNKVLVFDKNTGALTDTSOYQ	FKGLTKLNNDYTPHNOVNF	240
Qy	241	ENTSLETIDNYTAD	SWRPKDILKNGKWTWSSSDLR	PPLMSWMPDKQTQIAYLN	300
Db	241	ENTSLETIDNYTAD	SWRPKDILKNGKWTWSSSDLR	PPLMSWMPDKQTQIAYLN	300
Qy	301	QQGLGTGENYTAD	SSQESLNLAATQVQVKIETKI	ISQTQOTQWLRIINSFVK	TQPNWNSQ 360
Db	301	QQGLGTGENYTAD	SSQESLNLAATQVQVKIETKI	ISQTQOTQWLRIINSFVK	TQPNWNSQ 360

361 QY TSDTSAGEKDHLOGGALLYSNDSKTYANDSVRLNRTPTSTQTKPKYFEDNSSGGYDF 420
361 Db TSDTSAGEKDHLOGGALLYSNDSKTYANDSVRLNRTPTSTQTKPKYFEDNSSGGYDF 420
421 QY LLANDIDNSNPVQAEQLNWLHLYLMYGSIVANDPEANFDGVRVADVANNADLLQIASD 480
421 Db LLANDIDNSNPVQAEQLNWLHLYLMYGSIVANDPEANFDGVRVADVANNADLLQIASD 480
481 QY YLKAHYGVDSKSEKNAIHNLSILEAWSNDPQYNKDTKGALPDKLRLSLLYALTRPLE 540
481 Db YLKAHYGVDSKSEKNAIHNLSILEAWSNDPQYNKDTKGALPDKLRLSLLYALTRPLE 540
541 QY KDSASNEIRSGLEPVITNSLNRSAGKNSERMANYIFIRAHDSVQTVIAKIIKAQIN 600
541 Db KDSASNEIRSGLEPVITNSLNRSAGKNSERMANYIFIRAHDSVQTVIAKIIKAQIN 600
601 QY PKTDGTLFTLDELKQAFKYNEDMRQAKKYTQSNIPATYALMLSNKDSITRLYYGDMYS 660
601 Db PKTDGTLFTLDELKQAFKYNEDMRQAKKYTQSNIPATYALMLSNKDSITRLYYGDMYS 660
661 QY DGQYMATKSPYYDAIDTLLKARIKVAAGQDMKIITYVEGDKSHMDWDYTGVLTSVRYGT 720
661 Db DGQYMATKSPYYDAIDTLLKARIKVAAGQDMKIITYVEGDKSHMDWDYTGVLTSVRYGT 720
721 QY GANEATDQGESEATKTQGMVITSNPFLKLNQNDKVIIVNMGAAKHQOEYRPLILTTKDL 780
721 Db GANEATDQGESEATKTQGMVITSNPFLKLNQNDKVIIVNMGAAKHQOEYRPLILTTKDL 780
781 QY TSVTSDAAKSLYRKNDKGELVFDASDIQGYLNPQVSGYLAWVPVGSADNDQVRVAAS 840
781 Db TSVTSDAAKSLYRKNDKGELVFDASDIQGYLNPQVSGYLAWVPVGSADNDQVRVAAS 840
841 QY NKANATGQVYESSALDSQLIYEGFNFQDFVTKDSYTNKKIAQNVLFKSGNGVTSFEM 900
841 Db NKANATGQVYESSALDSQLIYEGFNFQDFVTKDSYTNKKIAQNVLFKSGNGVTSFEM 900
901 QY APOYVSEDSFLDSIIQNGYAFEDRYDLAMSNNKYGSQODMINAVKALHSGIOVIAD 960
901 Db APOYVSEDSFLDSIIQNGYAFEDRYDLAMSNNKYGSQODMINAVKALHSGIOVIAD 960
961 QY WVPDQIYNIPGKEVWTATRVNDGEYRKDSEIKNTLYAANTKSNKGDKYQAKYGGAFSEL 1020
961 Db WVPDQIYNIPGKEVWTATRVNDGEYRKDSEIKNTLYAANTKSNKGDKYQAKYGGAFSEL 1020
1021 QY AAKYPSIFNRTQISNGKKIDPSEKITAWKAKYPNGTNILGRGVYVLDKONASDKYFELKG 1080
1021 Db AAKYPSIFNRTQISNGKKIDPSEKITAWKAKYPNGTNILGRGVYVLDKONASDKYFELKG 1080
1081 QY NOTYLPKQMTNKEASTGFVNDGNGMTFYSTGYQAKNSFVODAKGNWYFDNNGHMYGL 1140
1081 Db NOTYLPKQMTNKEASTGFVNDGNGMTFYSTGYQAKNSFVODAKGNWYFDNNGHMYGL 1140
1141 QY QQLNGEVQVFLSNGVOLRESFLENADGSKNYFGLGNRSNGYYSFNDNSKRYFDASGV 1200
1141 Db QQLNGEVQVFLSNGVOLRESFLENADGSKNYFGLGNRSNGYYSFNDNSKRYFDASGV 1200
1201 QY MAVGLKTINGTQYFQDQGVQVKGAWITGSDGKKRYFDDGSGNMAVNRFANDKNGDWYYL 1260
1201 Db MAVGLKTINGTQYFQDQGVQVKGAWITGSDGKKRYFDDGSGNMAVNRFANDKNGDWYYL 1260
1261 QY NSDGIALVGVQTINGKTYTFQDQKQIKGKIITDNGKLYFLANSSELARNIPATDSQNN 1320
1261 Db NSDGIALVGVQTINGKTYTFQDQKQIKGKIITDNGKLYFLANSSELARNIPATDSQNN 1320
1321 QY WYFVSGDGVAVTGSQTIAGKLYFASDQGVKGSFVTVYNGKHYYHADSGELQVNRFEAD 1380
1321 Db WYFVSGDGVAVTGSQTIAGKLYFASDQGVKGSFVTVYNGKHYYHADSGELQVNRFEAD 1380
1381 QY KQGNWYFLDSNGBALTGSQRINDQVFFTRREGQVKGVDVAYDERRLLVY 1429
1381 Db KQGNWYFLDSNGBALTGSQRINDQVFFTRREGQVKGVDVAYDERRLLVY 1429

RESULT 2

B33135

gtfB protein precursor - Streptococcus mutans

C:Species: Streptococcus mutans

C:Date: 23-Oct-1990 #sequence_revision 23-Oct-1990 #text_change 09-Jul-2004

C:Accession: B33135; A33128

R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.

J. Bacteriol. 169, 4263-4270, 1987

A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.

A:Reference number: A33135; MUID:87308013; PMID:3040685

A:Accession: B33135

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1475 <SH1>

A:Cross-references: UNIPROT:P08987; UNIPARC:UPI000014D972; GB:M17361; NID:g153639; PIDN

R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.

submitted to the Protein Sequence Database, September 1990

A:Reference number: A33128

A:Accession: A33128

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-171,173-641,'N',643-1475 <SH2>

A:Cross-references: UNIPARC:UPI000017AC5E

A:Experimental source: strain GS-5

F:1096-1115/Domain: cpl repeat homology <CP1>

F:1224-1243/Domain: cpl repeat homology <CP2>

F:1289-1308/Domain: cpl repeat homology <CP3>

F:1354-1373/Domain: cpl repeat homology <CP4>

F:1419-1438/Domain: cpl repeat homology <CP5>

Query Match

Best Local Similarity 51.2%; Score 3817.5; DB 2; Length 1475;

Matches 762; Conservative 230; Mismatches 395; Indels 53; Gaps 22;

QY 1 METKRYKMHKVKHWTVAAGSLITLCTTTTGGSSVSAETBQOTSDKVVVTKSEDDKAA 60
Db 1 MDKVKYKLRKVKRWTVSVASAVMTL--TTLSSGLVKADSNESK-----QISNDNNTS 54
QY 61 SSSQTDAPKTQAOETQQAQSQAN--VADTSTSTTKETPPSQNTTQANSDDKTVNTKGS 119
Db 55 VVTANESNVIETATSKOEAASQTNHTVTTSSTSTSVVNVKVVSNPYTVGETASNGEK 114
QY 120 EBAQTEBETKQSEBAQTASSQALTQAKAELTKQRTAAQENKQPVDLAALPNVKQIDG 179
Db 115 LQNTQTTV--DKTSEAAANNISKQT--TEADTVDDSNAA----NLQILEKLPNVKEIDG 167
QY 180 KYIYIGSDGQPKNFALTVNNKVLVFDKNTGALTDTTS--QYQPKQGLTKLNNDYTPHNQIV 238
Db 168 KYIYYDNGKVKRTNFTLIADGKILHFDE--TGAYTDTSIDTVNKDIDVTTNSLYKKYQVY 226
QY 239 NFENTSLETIDNVYADSWYRPKDILKNGKWTWASSEDLRPLLLMSWPDKQTOIAYLNY 298
Db 227 DRSAQSFHFVDHYLTAESWYRPKYILKDGKTTQSTEKDFRPLMTWPDQETQRYVNY 286
QY 299 MNQOGLTGENTYADSSQESNLAAQTVQKLETKISQTOQWLRIINSFVKTPQNN 358
Db 287 MNAQ--LGINKTYDDTSNQLNIAAATQAKIEAKITLLKNTDMLGRQTSISAFVKTSQAMN 345
QY 359 SOTESDTSAGEKDHLOGGALLYSNDSK--TAYANSDYRLNRTPTSTQTK--PKYFEDNSS 415
Db 346 SDEKPFDD---DHLQNGAVLYDNEGKLTTPYANSVYRLNRTPTTQTKKDPRTADNTI 401
QY 416 GGYVFLANDIDNSNPVQAEQLNWLHLYLMYGSIVANDPEANFDGVRVADVANNADLL 475
Db 402 GGYVFLANDVDNSNPVQAEQLNWLHLYLMYGSIVANDPEANFDGVRVADVANNADLL 461
QY 476 QIASYLYKAHYGVDSKSEKNAIHNLSILEAWSNDPQYNKDTKGALPDKLRLSLLYAL 535
Db 462 QIAGDYLYKAAGIHKNDKAANDHLSILEAWSNDPTLYLHDDGNNMNMKNKRLSLFL 521
QY 536 TRPLEKDSNKNQNEIRSGLEPVITNSLNRSAGKNSERMANYIFIRAHDSVQTVIAKII 595

Db 522 AKPLNQ-----RSGMPLIINSLVNRDDEAETAAPSVSIFIRAHDSVQDLADI 573
QY 596 KAQINPKTDLTFTLDELKAFKIYNEDMRQAKKCYTQSNIPYAYALMLSNKDSITRLYY 655
Db 574 KAEINPNVGYSTMBEIKKAFIYNKDLATKEKKYTHYNTALSYALLTNKSSVRVYY 633
QY 656 GDMYSDGQWATKSPYDAIDTLLKARIKAYAGGQDMKITYYEGDKSHMDWDTYGVLT 715
Db 634 GDMFTDDGQWMAHTKTYNIEATLLKARIKAYSGQAMRNQOVGNSE-----IITS 684
QY 716 VRYGTGANEATDOGSSEATKTQGMVITSSNPISKLJNQNDKVIYVMGAHKNQYRPLILT 775
Db 685 VRYGKALKATDGTDRTRTSVAVIEGNPNURLKASDRVVMGAHKNQYRPLILT 744
QY 776 TKDGLSYTSDDAAKSLYRKNDELVFDASIQYLNQVSGYLAVVVPVGCASDNQDV 835
Db 745 TDNGIRKAYHSDQEAAGLVRYTNRGELIFTAAADIKYANPQVSGYLGVMVPVGA---LI 801
QY 836 RVAASHKANATQV--YESSALDSQLIYEGFNPDQVFTKDSYTNKKIAQNVQLFKSW 893
Db 802 KMFALRLARPHQQWASVHQNAALDSRVMEFGFSNFOAFATKKEEYTNVVIKNNVDFAEW 861
QY 894 GVTSEFMAPOYVSEDSGSLDSILONGYAFEDRYDLAMSKNNKYGSGQODMINAVKALHS 953
Db 862 GVTDFEMAPQYVSESDGSLDSVIQNGYAFETDRYDLGISKPNKYGTADDLVKAIALHSK 921
QY 954 GIOVIADWPDQIYNLPKGEVVTATRVNDYGEYRKDSEIKNTLYAAATKNSGKDYOAKYG 1013
Db 922 GIKVMADWPDMQYAFPEKEVVTATRVNDYGEYRKDSEIKNTLYAAATKNSGKDYOAKYG 981
QY 1014 GAFSELAQKYPISFNRTQISNGKKIDPSEKITAWKAKYPNGTNIILGRGVYVYLKONAS 1073
Db 982 GAFLELOAKYPPELFARKQISTGVPMDPSPVKIKQWSAKYFNGTNIILGRGAGYVYLKQATN 1041
QY 1074 KYPELKGQ--TYLPQMTNKEASTGFVNDGNGMTFYTSGYQAKNSFVQDACKNYYFD 1131
Db 1042 TYFNISDNKEINFLPRLTNLQDQVGSFYDGYVYTSYSGYQAKNTFISEG--DKWYYFD 1100
QY 1132 NNGHMYGLQQLNGEVOYFLSNQVQLRESFLENADGSKNYFGLHGNRYSGYSPFNDKSK 1191
Db 1101 NNGYMTGQASINGVNYFLSNGLQRLDAILKNEGTAYYAGNDRYENGYQF--MSGV 1159
QY 1192 WRYFDASGVNAGLKTINGNTQYFDQDGYQVKGAWITGDKKRYFDGSGMNAVRNFPAN 1251
Db 1160 WRHFN--NGEMSGLTVIDGQVQVDEMGYQAKGKFTVTDGKIRYFDKQSGMNYRNFIE 1218
QY 1252 DRNGDWYILNSDIALGVQTINGKTYYPQDQKQIKGLIITP--NGKLKYLANSGLAR 1310
Db 1219 NEEGKWLILGEDGAAVTGSQTINGQHLVFRANGVQVKGFEVTDPHHGRISYIDGNSGDQIR 1278
QY 1311 NIPATDSQNNWYFSGDGVAVTGSQTIAGKLYFASDGQVKGFSFVTYN--GKVHYHADS 1369
Db 1279 NRPVRNAQQWFFDNNGYAVTGARTINGQLLYFRANGVQVKGFEVTDYGRISYIDGNS 1338
QY 1370 GELQVNRFEADKGNWYILDSNGEALTGSRINDQVFFTRREGQVKGVDVADERLLVY 1429
Db 1339 GQOIRNRFRVNAQQWFFDNNGYAVTGARTINGQHLVFRANGVQVKGFEVTDHGRISY 1398
RESULT 3
T30858
glucosyltransferase - Streptococcus salivarius
C:Species: Streptococcus salivarius
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30858
R:Simpsen, C.L.; Giffard, P.M.; Jacques, N.A.
Infect. Immun. 63, 609-621, 1995
A:Title: Streptococcus salivarius ATCC 25975 possessses at least two genes coding for pri
A:Reference number: Z20909; MUID:95122197; PMID:7822030
A:Accession: T30858
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1577 <SIM>

A;Cross-references: UNIPROT:Q55265; UNIPARC:UPI000000B8087; EMBL:L35928; NID:G662380; P1
C;Genetics:
A;Gene: gtfm
Query Match 50.8%; Score 3788.5; DB 2; Length 1577;
Best Local Similarity 49.1%; Pred. No. 5.8e-158;
Matches 777; Conservative 208; Mismatches 412; Indels 185; Gaps 21;
QY 1 METKRYKXKHVKKXHWVTAVAS-----G 24
Db 1 MENKVPFKLHKXKXWVTIGVITLSVVALAGGSLLAQKVEADETSAPNGDGLQQLSEDG 60
QY 25 LITLGT-----LGSSVSAETBQOTSDKVVTKQKSEDDKAASSSSQTDAP 69
Db 61 TASLVTTTITVTEQASQAQASVAVATASVSHETSFOAATSQVSAQTAQAQTSFVASGEVA 120
QY 70 KYKQAT--EOTQAQSOANVADTSTISITKTPSQ----- 101
Db 121 VSSQTOSGQETQTTTEQVSGQSTQVAGQTSQAQTSFVTEQARPRVLTNAAPIATRAA 180
QY 102 -----NITTOANSDDKVT-----NTKSEEAQTSBEBTKQSEBAQTTASSQAL 144
Db 181 DSTIRINARNNTNITITAGTTPNVITITGPNTPKENVTVTSNGTRPNVTIIVTQPN 240
QY 145 TQAKAELTKQRQTAQAQENKPNVDLAAI--PNVKQIDGKYIYIGSDGQPKKNFALTNNKVL 203
Db 241 KPVPSPQSPQNPQVQNPQSLDYKPVASNLKTIIDGQIYV--ENGWVKKAAIELDGRLY 299
QY 204 YFDKNTGALTDSQYQFKOGLTKLNNDYTHQNIQNFENTSLETINYVTADSWYRPKOI 263
Db 300 YFDE--TGAMVDOSKPLRADAI PNNSIYAVYQAQYDTSKSFHDLNFLTADSWYRPKOI 358
QY 264 LKNGKKTWATASSDRLPRLMSWPDKOTQIAYLYNMNQGLGTGENVTADSSQESLNLA 323
Db 359 LKDGKNTASTKDRYPLMTWMPDKVTQVLYNTMSQQGFG--NKYTTTDMGSDLA 417
QY 324 QTVQVKIETKISQTOQTOWLRDIINSFVKTPQPNWNSQTESDTSAGEKDLHQQGALLYSNS 383
Db 418 ETVQGIIEERIGREGNTTWLRQLMSDFIKTPQGNWSESDNLLVG--KDHQQGALTFLNN 476
QY 384 DKTAVANSYRLNRTPTSGTKPKYFBNSSGQYDFLLANDIDNSNPVVOAQLNWLHY 443
Db 477 SATSHANSDFRLMNRTPNTQGTGRKYHIDRSNGGYELLANDIDNSNPVVOAQLNWLHY 536
QY 444 LMNYGSIIVANDPEANFDGVRVDVAVNNADLLQIASDYLKAHYGVDBKSEKNAIHLISILE 503
Db 537 IWNIGSILGNDPSANFDGVRIDAVNVDADLLQIASDYFKEKYRVADNEANAHLISILE 596
QY 504 AWSNDNDPQNKDTKGAQLPIDNKLRLSLLYALTPLPLEKDAKNKNEIRSGLEPVITNSLNN 563
Db 597 ANSYNDHQNKDTKGAQLSIDNPLRETLTTLFL-----KSNYRGSILERVITNSLNN 648
QY 564 RBAEGKNSERMANYIPIRAHDSVQTVIAKIKAQINPKTDGLTFTLDELKQAFKYNED 623
Db 649 RSSEQKHTPRDANYIPVRAHDSVQAVLANIISKQINPKTDGFTFTMDELKQAFEYNAD 708
QY 624 MEQAKKKTQSNIPYALWMLSNKDSITRLYYGDMYSDGQYMATKSPYYDAIDTLLKAR 683
Db 709 IAKDKKTYQYNIPIAATMLTNKDSITRYYIGDLFTDDGQYMAEKSPYNAIDALLRAR 768
QY 684 IKYAAAGGQDMKITTYEGDKSHMDWDYTVGLTSVRYGTGANEATDOGSSEATKTQGMVITSS 743
Db 769 IKYVAGGQDMKITVYKNGYE-----IMSSVRYGKGAEANQLGTAETRNQGLMLVITA 819
QY 744 NNPSILKLNQNDKVIYVMGAHKNQYRPLILTTLTKDGLTSYTSDA--AAKSLYRKNQDKGEL 802
Db 820 NRPDMKLGANDRLVVMGAHKNQYRPLILTSLKSTGLATYTKDSVDPAGLVRYTDMQGNL 879
QY 803 VFDASDIQGLNLPQYSGYLAVVVPVGCASDNQVVRVAASNKANATGVYVSSSSALDSOLIY 862
Db 880 TFDADDIAGHSTVEVSGYLAVVVPVGCASDNQARTKASSTKKGE--QVFSSAALDSQVIY 938
QY 863 EGFNSFQDFVTKDSYTNKKIAQNVQLFKSWGVTSPFMAPQYVSSSDGSLDSIIQNGVA 922

Db	939	EGFSNFQDFVKTPSQYTNRVIAQNAKLFKEWGITSEFPAPQVSSQDGFELSLIENGVA	998
Qy	923	FEDRYDLAMS KNNKYQSQQDMINAVKALHKSIGQVIADWVPDQIYNLPKQEVVATRVND	982
Db	999	FEDRYDIAMS KNNKYGSLKDLMDALRALHAEGISAIADWVPDQIYNLPKQEVVATSRNVS	1058
Qy	983	YGEYRKDSSEIKNTLYAANTKSNKGYQAKYGGAFISELAAPQSIINFRQISNGKKIDPS	1042
Db	1059	YGTPEPNAEYINSLYAARTFTGDNDFQGGYGGAFDELKAKYPAIFERVQISNGRKLTTN	1118
Qy	1043	EKITAMKAKYFNGTINILGRGVYLVKDNASDKYFELKGNQTYLPKQMTNKEASTGVPNDG	1102
Db	1119	EKITQSAKYFNGSNTQGTGARVYLDQNAATNQYFSVKAQGTFLPKQMTITGS-GRFRVG	1177
Qy	1103	NGMTFYSTSGYQAKNSFVQDAKGNWYYPDNNGHMYGLQQLNGEVQYFLSNGVQLRESFL	1162
Db	1178	DDVQYLSIGYLAKNTFIQVGANOWYYPDKNGNMVTGEQVIDGKYFFLDNGLQLRHVLR	1237
Qy	1163	ENADGSKNYPFGLHGRNYSNGYISFPDNDK-WRYFDSAGVMAVGLKTINGNTQYFDQ-DGY	1220
Db	1238	QGSDDGHVYYPDKGVOAFNGFYDFAGPRQDVRYFDGNGQMYRGLHDMYGTTFYFDEKGTI	1297
Qy	1221	QVKGAWITSGDKRYFDDGSGNMAVNEFA-NDKNGDWYLLNSDGLALGVQVOTINGKTY	1279
Db	1298	QAKKEIRFADGRTIRIFIDTGNLAVNRFAQNPENKAWYILSDNGYAVTGLQTINGKQY	1357
Qy	1280	FGQDGKQIKIITDNGKLYFL-ANSBELARINIFATDSQ-----	1318
Db	1358	FDNEGROVKGHFTVINNQ-RYFLDGSDEIAPSRFVTENKNWYVDGNGKLVKAQVING	1416
Qy	1319	-----NNWYFSGDGVAVTGSQTIAG	1339
Db	1417	NHYTFNNDYSQVKGAWANGRYDDGSGQAVSNQFIQIAANQWYLLNQDGHKVTGLQINN	1476
Qy	1340	KKLYFASDCKOVKGSFVYNGKVHYHADSGELQVNRFEADKGNWYLLDSNGEALTGSQ	1399
Db	1477	KVYFSGNAGVAKGKULTVQGGKCYCFDAHTGEQVNRFEAARGCWYFNSAGQAVTGOQ	1536
Qy	1400	RINDQRVFFTRREGQVKGVGVAY	1421
Db	1537	VINGKQLYFDGSGRQVKGKRVY	1558
RESULT 4			
T30857			
glucosyltransferase I- Streptococcus salivarius			
C:Species: Streptococcus salivarius			
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004			
C:Accession: T30857			
R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.			
A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for pri			
Infect. Immun. 63, 609-621, 1995			
A:Reference number: Z20909; MUID:95122197; PMID:7822030			
A:Accession: T30857			
A:Status: preliminary; translated from GB/EMBL/DBJ			
A:Molecule type: DNA			
A:Residues: 1-1449 <SIM>			
A:Cross-references: UNIPROT:Q55264; UNIPARC:UPI000008166B; EMBL:L35495; NID:g662378; PII			
C:Genetics:			
A:Gene: gtfL			
Query Match 49.8%; Score 3713.5; DB 2; Length 1449;			
Best Local Similarity 51.4%; Pred. No. 9.8e-155;			
Matches 754; Conservative 207; Mismatches 409; Indels 97; Gaps 22;			
Qy	1	METKRYKQHKVKKHVTAVV-----ASGLITLGT-----TTLGSSVSAET	41
Db	1	MDKKVHYKHKVKQWVTAVTGLSLGAVSAVSLGTGNDGVQVQADEHTATVAIPDITVDI	60
Qy	42	EQOTSQKVVTKSEDDKAASESQTD--APKTKQAQTEQTQAQSQANVADTSTSI-----94	
Db	61	GTVSNUTTAQAQDPTTAVATNDVATQATPTATFDLTDTTNTVAANAVDTVATVGTTRA	120

Qy	95	--TKETPSONIT--QANSDDKTVTNKSEEAQTSERTKQSEAAQTASSOALTOAKAE	150
Db	121	ATTNDTATNDTAVDTTNNNTTDTTVDTRAATTERRATGARRGT--GGRATTVPNGN	178
Qy	151	LTKQRTAAQENKNPVDLAAIPNV-----KQIDGKYIYIGSDGQPKNFALTVNNKV	202
Db	179	TNNANTVTVNN--DLPATNNVTDGSPSHIKTINGQYVVEDDGTIRKNVLERIGS	235
Qy	203	LYFDKNTGALTPTSQYQF-KOGLTKLNNDYTPHNOIVN-----FENTS---LETIDNYVT	253
Db	236	QYFNAETGELSKEQYRFDKNGCTGSSADSTNTVTVNGDKNAFYGTDTDKDELVDGYET	295
Qy	254	ADSWRPDKDILKNGTKWTATASSEDRLPILMSWMPDKQTOIAYLVNWNQOGLGTGENYTAD	313
Db	296	ANTWYRPKEILKDGKEWTASTENDKRPLLTVMWPSKAIQASLYNMYKEGLGTNGTYTTF	355
Qy	314	SQESINLAAQTVQVKIETKISQTOQWLRDIINSFVKTPQPNWNSQTESDTSAGEKHL	373
Db	356	SSQTDMDQALLEVQKRIERIAREGNTDMLRTTIKNFVKTPQWNSTSE---NLDNNDHL	412
Qy	374	QGGALLYNSDKTAYANGSYRLNLRTPPTSQTK--PKYPEDNSSGGYDFLLANDIDNSNP	431
Db	413	QGGALLXNDSRTSHANSYRLNLRTPPTSQTKHNPKYTKDTSNGGFEFLANDIDNSNP	472
Qy	432	VVOAQELNWLHLMYGYSIVANDPEANPDGVRDAVDVNVNADLLQIASYLYKAHYGVDS	491
Db	473	AVQAEQLNWLHYIMNIGTITGGSDEDFDGVDAVDVNVNADLLQIASYLYKAHYGVDS	532
Qy	492	EKNALNHLISLEAWSNDPQYNKDTKGAOLPIDNKLRLSLLYALTPLPKDASNKNEIRS	551
Db	533	QQQALKHLISLEAWSNDAYNEDTKGAOLPMDPMLALVLSLLRPIGN-----RS	584
Qy	552	GLEPVTITSLNRSABGKNSERMANYIFIRAHDSVQTVIAKIIKAQINPKTDGLTFTLD	611
Db	585	GVEPLISLNSLDSRSESGKNSKRMANYAFVRAHDSVQSIIGLIKNEINPQSTGNTFTLD	644
Qy	612	ELKQAFKIYNEDMROAKKYITQSNTPTAYALMNSKDSITRLYGYDMYSDGQYMATKSP	671
Db	645	EMKKAPEIYNKDMRSANKQYTOYNIPSAYALMLTHKDTVPVYGYDMYTDGQYMAQKSP	704
Qy	672	YYDAJDTLLKARIKYAAGGQDMKITVVEGDKSHMDWDYTGVLTSVRYGTGANEATQGSSE	731
Db	705	YYDALETLLKGRIRVYAGGQDMKNVYI-GYGTNGWDAAAGVLTSVRYGTGANSASDTGTA	763
Qy	732	ATKTQGMVITSNPNSLKLNDQNVIVNMGAAHQEQYRPLLLTTKDGITSYTSDDAAKS	791
Db	764	ETRNQGMVIVSNQPALRLTSN--LTINMGAAHRNQAYRPLLLTTNDGVATVYLNDSANG	821
Qy	792	LYRKTNDKGELVFDASDIQYLNPOVSGYLAVWVPVGASDNODVRAASNKANATQVYVE	851
Db	822	IVKYTDGNGNLTFSAINEIRGNPQVDGYLAVWVPVGASENQDVRVAPSKENSSGLVYE	881
Qy	852	SSSALDSQLIYRFGSNFQDFTKSDYTNKKAQNVQLFKSWGTSTFEMAPQVSSSEDS	911
Db	882	SNAALDSQVIYRFGSNFQDQVQNPQYTNKKIAENANLFKSWGITSFEFAPQVYVSSDDGS	941
Qy	912	FLDSIQNGYAFEDRYDLAMS KNNKYGSQQDMINAVKALHKSIGQVIADWVPDQIYNLP	971
Db	942	FLDSVIQNGYAFTRYDIQMSKDNKYGLADLKAALKSLHAVGISAIADWVPDQIYNLP	1001
Qy	972	KEWVTRVNDYGEYRKDSSEIKNTLYAANTKSNKGYQAKYGGAFISELAAPQSIINFR	1031
Db	1002	DEVVTRVNNYGETKDGAIIDHLSYAAKTRTFPGNDYQYGGAFDELKRLYPPQIFDRV	1061
Qy	1032	QISNGKKIDPSEKITAWKAKYFNGTINILGRGVYLVKDNASDKYFELKGNQTYLPKQMTN	1091
Db	1062	QISTGKRWTTDEKITQWSAKYNGTINILDRGSEYVLK-NGLNGYYGTNGKGYSLPKVGS	1120
Qy	1092	KEASTGTFVNDGNG-----MTFYSTSGYQAKNSFVQDAKGNWYYPDNNGHMYGLO	1141
Db	1121	NQSTNGDNQNGDGSKGKFLRFSVRYRNNQGYAKNAFIKDNNGVYVFDNSGRMAVGEK	1180

A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1592 <ABO>
A;Cross-references: UNIPARC:UPI00012BCB2; GB:D90213; NID:g217032; PIDN:BAAL4241.1; PID:
F;1093-1112/Domain: cpl repeat homology <CP1>
F;1222-1241/Domain: cpl repeat homology <CP2>
F;1287-1306/Domain: cpl repeat homology <CP3>
F;1330-1351/Domain: cpl repeat homology <CP4>
F;1352-1371/Domain: cpl repeat homology <CP5>
F;1402-1420/Domain: cpl repeat homology <CP6>
F;1465-1484/Domain: cpl repeat homology <CP7>
F;1513-1532/Domain: cpl repeat homology <CP8>

Query Match 47.5%; Score 3542; DB 2; Length 1592;
Best Local Similarity 48.0%; Pred. No. 3.6e-147;
Matches 714; Conservative 242; Mismatches 411; Indels 120; Gaps 27;

QY 1 METKRYKMKHKVKKHVVAVASGLITLGTITLGGSSVSAETEQTSDKVVTKSEDDKAA 60
DB 1 MEKNVRFKMKHKVKKRWVTLVSASA--TMLASALGASVAS-----ADTDAS 44

QY 61 SESSQTDAPKTKQAQTEQTAQQAQANVADTSTSTTETPSONITTOANSDDKTVTKSE 120
DB 45 DDSNQ-----VVTGDTNNQATQTSIAATATS--EQASTDAATDQASAEQQTGTFAT 100

QY 121 EAQTSBERTKQSEBAQ---TTASSQALTOAKAELTKQRTAAQENKNPVDLAAIPNVKQI 177
DB 101 D--TAAQTITNANEAKWVPTENENQGTDEMLAEAKNVAT--AESDSIPSLAKMSNVKQV 157

QY 178 DGKYYIGSDGQPKKFPALTVNNKLVYDFKNTGALTDTTSQYQFQKGLTKLND---YTPH 234
DB 158 DGKYYIQDQGNVKKNFVAVSGVKIYYFDB--TGAYKDTSDVDADKSSVASQATIFAAN 216

QY 235 NQIVNFENTSLETIDNVVTADSVWRPKDILKNKGTWTASSEDLRLPLMSWPKDQTOIA 294
DB 217 NRAYSTSAKNFEADVNYLTRADSVWRPKSILKQKWTESGDDFRPLMAWVDETTETKRN 276

QY 295 YLYNMNQGLGTGENTYADSSQBSLNLAAQTVQVKIETKISQQTQWLRIIDINSFVKQ 354
DB 277 YVNMNKKV--VGIDKTYTAETSQADLTAAAEVLQARIEQKITSENNTKWLREALSAFVKQ 335

QY 355 PNWNSPTESDTSAGEKHQGGALLYSN--SDKTAYANSYRLLNRTPTSGT-----KPKY 409
DB 336 PQWNGSESEKPYD----DHLQNGALLFDNQDLDLTPDQSNRYLLNRTPTNQGLSDSRFTY 391

QY 410 FEDNSGGYDFLLANDIDNSNPVVAQELNWLHLNMGSIIVANDPEANFDGVRVDAVN 469
DB 392 NPNDPLGGYDFLLANDVDNSNPVVAQELNWLHLNMGSIIVANDADANFDSIRVDAEN 451

QY 470 VNADLLQIASDYLKAHYGVDKSEKNAIHLNLSLEAWSNDNDPQYNKDTKGAQPLDNKJRL 529
DB 452 VDADQLQISSDYLKAAYGIDKNNKNANNHVSIVAEWSNDNDTPYLHDDGDNLMNMNKFRL 511

QY 530 SLLYALTRPLEKDAKSNKIRSGLEPVTNSLNRRSAEKNRMANIYIFIRAHDSVQT 589
DB 512 SMLWSLAKP-----TDVRSGLNPLIHNLDREVDDREVEVTPSYSFARAHDSVQD 563

QY 590 VTAKIILKAQINPKTDCGLTTLDELKQAFKLYNEDMPQAKKYTQSNIPATAYALMSNKDS 649
DB 564 IIRDIILKABINPNSFGYSTQBEIDQAFKIYNEDLKKSDKKYTHYNVPLSYTLTLNKG 623

QY 650 ITRLXYGDMYSDDGQYMATKSPYDAIDTLTKARIKYAAGGQDMKITTYVEGDKSHMDWDY 709
DB 624 IPRVYVYGMFTDDGQYMANKYVNYDAIESLLKARMYVAGQAMQ--NYQINGE----- 676

QY 710 TGVLTSVRYGTGANENATDQSEATKTOGMNAVITSNPNLSKLQNDKVIYVMGAHKNQBY 769
DB 677 --ILTSVRYGKALKQSDKGDATTRTSGVGVVGNQPNFSLD--GKVVALNMGAAHANQBY 733

QY 770 RPLLLTTKGLTSYTSDDAAK--SLYRKTNDKGELVFDASDITQGLYNPQVSGYLAWVWY 827
DB 734 RALMVSTKDGVAITYATDADASKAGLVKRTDENGYLEYFLNDDLKGVANPQVSGFLQVWV 793

QY 828 GASDNQDVRVAASNKANATQVYESSSALDSQLIYEGFSNFQDFVTTKDSYTNKKIAQNV 887
DB 794 GAADDQDIRVAASDSTASTDCSKLSLHQDAAMDSRVMEFGFSNFQSPATKEEYTVNVIANNV 853

QY 888 QLFKSGVTSFEMAPQVYSSSEDSFLDSIIIONGYAFEDRYDLAMSKNKYGSSQODMTNAV 947
DB 854 DKFVSWGITDFEMAPQVYSSSTDQGFQDSVIQNGYAFTRDYDLGMSKANKYGTADQLKAI 913

QY 948 KALHKSQIOVADWPDQIYNLPKQEVVTRVNDVGEYRKDSIKNTLVAANTKSGKD 1007
DB 914 KALHAKGLKVMADWVDPDMYTFPKQEVVTRTRDKFGKPIAGSQINHSIVTDTKSSGDD 973

QY 1008 YQAKYGAFLSELAAYPSIFNRTOISNGKKIDPSEKITAWKAKYFNGTILGRGVGYL 1067
DB 974 YQAKYGAFLDELKEKYPPELFTKKQWSTGQAIIDPSVKIKQWSAKYFNGSNILGRGADYVL 1033

QY 1068 KDNASKYFELKNGQYTLPKQMTNKEASTQFVNDGNGMTFYST--SGVQAKNSFVQDAKGN 1126
DB 1034 SDQVSNKYFNVASDITLFLPSSLGLGVVSGIRYDGGKIYINSSATGDQVKASFITEA--GN 1092

QY 1127 WYFDMNNGHMYVGLQQLNGERVQYFLSNGVOLRESFLENADGSKNYFGLGNRY--SNGYY 1184
DB 1093 LYTFGKDGWTVTGAQTINGANYFFLENGTALRNTIYTDAGNSHYVANDGRKYENENGYQ 1152

QY 1185 SPDNDKRWYFDASGVMAVGLKTINGTOYFDODGYQVKGAWITGSDGKRYFDDGSGNM 1244
DB 1153 QFGND--WRYF--KQGNMAGLTTVDGNVQYFDKDGQVQAKDKIIVTRDGKRYVFDHNGNA 1209

QY 1245 AVNRPANDKNGDWYILNSDGIALLVGVTQTINGKTYTYPGQDQKQIKGKIIT--DNOKLKYFLA 1303
DB 1210 VTNFTIADTKGHWYILGKGVAVTGAQTVGKQKLYFEANGQQVKGDFVTSDEGLKLYFDV 1269

QY 1304 NSGELARNIPATDSQNNWYFSGSDGVAVTGSQTIAGKLLYFASDGKQVGSFTV--YNGKV 1362
DB 1270 DSGDMWTFIEDKAGNWFYLGKGAATVGAQTIKQKLYFRANGQQVKGDIIVKGTGDKI 1329

QY 1363 HYHADSGELQVNRFEADKDG----- 1383
DB 1330 RYVDKASGQVFNKTVKAAADGKTYVIGNDGVADVPVVKVQTFPKDASGALRFYNLKGQIV 1389

QY 1384 -----NWYILDSNGEALTGSRQINDQVRVFTREGKQVKGDV 1419
DB 1390 TSGMWYETANHDWVYIQS--GKALTGSEQTINGOHLFYFKDGHQVKGQL 1435

RESULT 8
A44811
glucosyltransferase (EC 2.4.1.-) I - Streptococcus salivarius
C;Species: Streptococcus salivarius
C;Date: 31-Dec-1993 #sequence revision 31-Dec-1993 #text_change 09-Jul-2004
C;Accession: A44811; S27276; S28809
R;Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.
J. Gen. Microbiol. 137, 2577-2593, 1991
A;Title: Molecular characterization of a cluster of at least two glucosyltransferase ge
A;Reference number: A44811; MUID:92148377; PMID:1838391
A;Accession: A44811
A;Molecule type: DNA
A;Residues: 1-1518 <GIF>
A;Cross-references: UNIPROT:Q00600; UNIPARC:UPI000008BEF31; EMBL:Z11873; NID:g47526; PID
A;Note: sequence extracted from NCBI backbone (NCBI:81050, NCBI:81052)
C;Genetics:
A;Gene: gtfJ
C;Keywords: glucosyltransferase; hexosyltransferase
P;1307-1326/Domain: cpl repeat homology <CP4>

Query Match 46.9%; Score 3502.5; DB 2; Length 1518;
Best Local Similarity 48.1%; Pred. No. 1.8e-145;
Matches 724; Conservative 247; Mismatches 434; Indels 101; Gaps 32;

QY 1 METKRYKMKHKVKKHVVAVAS--GLITL--GTTILGSSVSAETEQTSDKVVTKSEDD 57
DB 1 MENKHYKLHKVKKQWTVIAVASVALATVGLGLSVTTSSVSAD---ETQDKTVTQ--SNSG 56

QY 58 KAASESQTDAKPKTQAQTEQTOAQ-----SQANVADTSTSIKETPSONITTOANSDDK 112
DB 57 TTASLVTSPSEATKEADKRTNKTEADVLTPAKETNAVETATTTNTQATAEAAT-ATTADV 115
QY 113 TVTNTKSEAAQTSEE---RTKQSEBAQTASSQAL-TQAKAELTKQRTAAQENKNPVD 167
DB 116 AVAAVNPKAEVTTDAPAVTTTEKAEQOPATVRAEVNTEVKA-----PENALKDSEVEA 169
QY 168 LAAIPNVKQIDGKYIYISDGQPKKNFALTVNNKVLVFDKNTGALTDTTQYQFKQGLTKL 227
DB 170 ALSLKNIKNIDGKYIYVNEGSHKENFAITVNGQLLYFGKD-GALTSSTYSPTPTWTNI 228
QY 228 NNDYTHNQINVPENTSLLETIDNYVTADSWYRPKDLKNGKWTWASSESDRLPLMSWHP 287
DB 229 VDGFSINNRAYDSSEASFELIDGLYTADSWYRPASIIKDGVTWQASTAEDFRPLMAWHP 288
QY 288 DKOTQIAYLNNMOGLTGENTYADSSOESLNLAQAQTVQVKIETKISOTOOTOWLRDII 347
DB 289 NVDTQVNYLNNYSKV-FNLDAKYSSTDQKETLKVAAKDQIKLEQKIQAEKSTQWLRETI 347
QY 348 NSFVKTPQNNWSQTESDTSAGEKDHLOGGALLYNSNDKTAYANSDYRLNRTPTSGTK- 406
DB 348 SAFVKTPQNNKETENYSKGGEDHLQGGALLYVNDSTRTPWANSDYRLNRTATNQTGI 407
QY 407 -----PKYPEDNSSGGYDFLLANDIDNSNPVQAEQOLNMLHYLMNYGSIIVANDPEANFDG 461
DB 408 DKSILDEQSDPNHMGDFLLANDVLSNPVQAEQOLNQHLYLMNNGSIVMGDKDANFDG 467
QY 462 VRVDVANNVADLLOJASDYLKAHYGVDKSEKNAIINHLSILEAWSNDNDPOYNKDKTGAOL 521
DB 468 IRVDADVNDADMLQYITVFRYRYGVNKSANALAHISVLEAWSLNDHNDKTDGAAAL 527
QY 522 PIDNKLRLSLLYALTRPLEKDSANKNEIRSGLEPVTNSLN----- 562
DB 528 AMENKQRLALFSLAKPIK-----RTPAVSPLVNNTFNTQRDEKTDWINKOGSKAY 580
QY 563 -----NRBAEGKNSE-----MANYIFIRAHSEVQTVIAKIIKAOINPKTDGLTFLDE 612
DB 581 NEDGTVKQSTIGRYNEKYGDASGNYVFIRAHDNNVQDIIAEIIKBINPKSDGFTTDAE 640
QY 613 LKQAFKIYNEDMROAKKYTOSNIPYAYALMLSNKDSITRLYYGDMVSDDGQYMAKSPY 672
DB 641 MKQAFIYKNDMLSSDKYTLNINIPAYAVMLQMETITRYYGDIYTDGHHMETKSPY 700
QY 673 YDAIDTLKARIKYAAGQDKMITYV---EGDKSHMD---WDYTGVLTSVRYGTGANEATD 727
DB 701 YDTIVNLMKSRIKYVSGQAQRSYWLPDGTGKMDNSDVELYRTNEVYTSVRYGKDIWAND 760
QY 728 -QGSEATKTCGMAVITSNPSSLKLNQNDKVIYVNGAAHKNQOEYRPLLLTTKDLTSYTS- 785
DB 761 TEGSKYSRTSGQVTLVANNPKLNDQSAKLNVEMGKIHANQYRALIVGTADGIRKFTSD 820
QY 786 -DAAAKSLYRKTNDKELVFDASDIOCYLNPQVSGYLAVMVPGASDNQDVRVAASNKAN 844
DB 821 ADIAAGYKETSUNGVLIFGANDIIKGYETFDMSGFVAVVWVPGASDNQDIRVAPSTEAK 880
QY 845 ATGOV-YESSALDSQLIYEGFSNFQDFVTKDSD---YTNKKIAQNVLFKSNGVTSFEM 900
DB 881 KEGELTLKATEAYDSQLIYEGFSNFQT-IPDGSDDPSVYTNRKIAENVDLFKSNGVTSFEM 939
QY 901 APOYVSESDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSQODMINAVKALHKSGIQVIAD 960
DB 940 APOFVSADGDTFLDSVIQNGYAFADRYDLAMSKNNKYGSKEDLRDALKALHKAGIOAIAD 999
QY 961 WVPDQLYNLPKGEVWVATRVNDYGEYRKDSEIKNTLYAANTKSNKGDKYOAKYGGAFISEL 1020
DB 1000 WVPDQIYQLPCKEVVATRTDAGRKRIADAIIDHSLYIVANSKSGDKYOAKYGGEFIAEL 1059
QY 1021 AAKYPSIFNRTQISNGKIDIPSEBKI-TAWKAKYFNGTNILGRGVGYVLKONASDKYFELKG 1080
DB 1060 KAKYPMFKVNMISTGKPIDDSVKLKQWKAERYFNGTNVLERGVGYVLSDEATCKYFTVTK 1119
QY 1081 NQTYLPKQMTNKE-ASTGPFVNDNGMTFFYSTSGYQAKNSFVQDAKGNWYFFDNNGHVYG 1139

DB 1120 EGNFIPQLUTGKEKUITGFSSDGKGITFTGTSGTQAKSAFV-TFNGNTYFFDARGHMVTN 1178
QY 1140 LQ-QLNG-EVOYFLSNGVOLRESFLENADGSKNYFEGHLGNRYNSNGYISFDNDS----- 1190
DB 1179 SEVSPNGKDVYRFLNGIMLSNAFVIDANGNTLYNSKGOMYKGGYTKFEDVSETDKDGE 1238
QY 1191 ----KWRFPDASGVMAVGLKTINGTQYFPDQGYQVKGAWITGSDGKKRYFPDGSQNAV 1246
DB 1239 SKVWFRFYFTNBGVMAKGVTVIDGFTQYFGEDGFOAKDLVT-FKGKTYFFDAHTGNGIK 1297
QY 1247 NRPANDKNGDMWYLSNDSGIALGVOTINGKTYVYFGDGKQIKGKIITD-NGKLKYFLANS 1305
DB 1298 DTWRN-INGKWTYFDANGVAATGAQVINGQKLYFNEGDSQVKGGVVKNADGYTSKYKEGF 1356
QY 1306 GELARNIFATDSQNNWYFPGSDGVAVTGSGQTIAKKLYPASDGKQVKGSGFV-TYNGKVHY 1364
DB 1357 GELVTFNEFTDGNWYVYAGANGKVTGTAQVINGQHLVFNADGSOVKGGVVKNADGYTSK 1416
QY 1365 YHADSGLQVNRPEADKQGNWYVYLDNSGEALTGSORINDQVFFFTREGKQVKGK-D-VAYDE 1423
DB 1417 YNASTGERLITNEFTTGDNNWYIIGANGKSVTGEVKIGDDTYFFFAKDGKQVKGQTVSAGN 1476
QY 1424 RELLVY 1429
DB 1477 GRISY 1482

RESULT 9
S22737
glucosyltransferase (EC 2.4.1.-) S - Streptococcus salivarius
C:Species: Streptococcus salivarius
C:Date: 16-Sep-1992 #sequence revision 16-Sep-1992 #text_change 09-Jul-2004
C:Accession: S22737; S28810; B44811; S22727
R:Jacques, N.
submitted to the EMBL Data Library, March 1992
A:Reference number: S22726
A:Accession: S22737
A:Molecule type: DNA
A:Residues: 1-1599 <JAC>
A:Cross-references: UNIPROT:Q00599; UNIPARC:UPI000000BEF34; EMBL:Z11872; MID:g47530; PIDN
A:Experimental source: ATCC 25975
R:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.
J. Gen. Microbiol. 137, 2577-2593, 1991
A:Title: Molecular characterization of a cluster of at least two glucosyltransferase ge
A:Reference number: A44811; MUID:92148377; PMID:1838391
A:Accession: S28810
A:Molecule type: DNA
A:Residues: 1-51 <GIF>
A:Cross-references: UNIPARC:UPI000017027C; EMBL:Z11873
C:Genetics:
A:Gene: gtfK
C:Keywords: glycosyltransferase; hexosyltransferase
F:1456-1475/Domain: cpl repeat homology <CPR>

Query Match 44.4%; Score 3312; DB 2; Length 1599;
Best Local Similarity 45.3%; Pred. No. 4.2e-137;
Matches 720; Conservative 202; Mismatches 466; Indels 200; Gaps 30;

QY 1 METKRYKMKHKVKKHWTVAVAS-GLITL--GTTTLGSSVSA-ETEQQTSDKVVTQKSED 56
DB 1 MENKRYKLUHKVKKQWTVLAVASVALATIVGSVATSSLASAEETNNSGSPSTTVGEN 60
QY 57 DKAASE---SSQTDAPKTKQAQTEQTAQSQANVADTSTSIKETPSQ-----NITQAN 108
DB 61 TNPVVEKEVGTTEVTANTSNATTTERAEVTADKPAGTTPQNSGTTSDRAAAVEVEAKPE 120
QY 109 SDDKTVNTNKSEAAQTSEERTKQSEEAQTASSQALTOAKAELTKQRTAAQENKNPVDL 168
DB 121 TTAKPEVATKPTATTSEVAANAGVAAPTTEKSKELSE-----EIKAAVSL 167
QY 169 AAI PNVKQIDGKYIYISDGQPKKNFALTVNNKVLVFDKNTGALTDTSTQYQFKQGLTKLN 228

Db 168 DN1K--KEKDKYKYLLEDGSHKQNFALTNGQVLYFDEN-GALSSSTSYSTPQETNNLV 224
QY 229 NDYTPHQNIVNFENTSLTIDNYVTADSWYRPKDILKNGKTTWTASSSDLRPLLMWPD 288
Db 225 TDFTKNAAVDSTKASPELVGYLTADSWYRPKEILLEAGTWTAKSTEKDFRPLLMWPD 284
QY 289 KQTQIAYLNTNMOGLGTGEN----YTADSSQBSLNLAAQTQVVKIETKISQTOQTOWLR 344
Db 285 KDTQVAYLNTMT-KALNSGBETKDVFTIENSQASLAAQAQILQKIEVKIAANKSTDWLR 343
QY 345 DIINSFVKTOPNNSOTESDTSAGEKDLGGALLYSNDSKTAAYANSDEYLLNRTPTSQT 404
Db 344 QSTEAFFVKDQKNWINSSE-----GKEHFGKALLFVNSDSTKANSYRKLNQATSYI 399
QY 405 GKPKYFEDSSGGYDFELLANDIDNSNPVQAEQLNMLHYLMNYGSIV--ANDPEANPDGV 462
Db 400 KNHK-IWNGSDGGEYFLLSNDIDNSNPVQAEMLNQLYFPMWQIVFGDKDKDAHFDGI 458
QY 463 RVDADVNADLLQIASDYLKAHYGVDKSEKNAINHLSILEANSNDNDPQNKDKTKGAQLP 522
Db 459 RVDADVNVSVDMLQLVSSYMKAAKYVNESEARALANISILEANSNDNDPYVNEHNTAALS 518
QY 523 IDNKLRLSLLYALTRPL-----EKDASNKNEIRSGLEPVTINSLNRSAGEKNSERM-- 574
Db 519 MDNGLRLSVHGLTRPVTNKGTRGARNASMKDLINGGY-----FGLSNR-AEVTSDQLGF 572
QY 575 ANYIFIRAHDSVQTVIAKIIKAQINPKTDGLTFTLDELKQAFKIYNEDMRQAKKYTOS 634
Db 573 ATYLFVRANDSEVQTVIADIISKIDPTDGTFTLDQLKQAFDIYNADMLKYDKKYTHS 632
QY 635 NIPTAVALMSKDSITRLYGYDMYSDGQYMAKSPYDAIDTLKARIKYAAGQDMK 694
Db 633 NIPAAVALMLQTMGAATRVYGLYTDNGQYMAKSPYFQDITLLKARPKYVAGQTSY 692
QY 695 ITTVEGDKSHMDWYTGVLTSVRYGTGANEATD-QGSEATKQGMVAITSNPSKLQN 753
Db 693 INHLADGVSSAKDNKEVLVSRYGQDLMSKTDTEGGYGRNSGMLTLIANNPDLKADG 752
QY 754 DKVIVNMGAAHKNQYRPLLLTTKGLTSTYSDAAAKSLYRKNDXGELVFOASDIQGYL 813
Db 753 EITIVNMGAAHKNQYRPLLLTTEKIGIVSLNDSDTK-IVKYTDAQGNLVFTADEIKGFK 811
QY 814 NPQVSGYLAVWPVPGASDNQDVRVAASNKANATG-QVYESSALDSQLIYEGFSNQDFV 872
Db 812 TVDMSGYLSVWPVPGATDQNVLAKPSTKAYKEGDKVYSSAALAEQAQVIEGFSNQDFV 871
QY 873 TKDSYTNKIAQNVQLFKSNGVTSSEMAPQVYSSDEGSLDSIIQNGYAFEDRYDLAMS 932
Db 872 KEDSQYTNKLIAANADLFKSWGITSFEIAPQVYSSKDGTFLLDSIIENGYAFTDRYDFAMS 931
QY 933 KNNKYSQODMTINAKALHKSIGIQTADWVPQIYNLPKGEVVTATRVNDYGEYRKDSFI 992
Db 932 KNNKYSKEDLRDALKALHKGIIQVADWVPQIYNLPKGEVVTATRVNDYGEYRKDSFI 991
QY 993 KNTLYAANTKSGKQYQAKYGAFSLBELAAKYPISFNRTQISNGKKIDPSEKITAWKAY 1052
Db 992 VNKLXYNTKSSGNDFOAQYGAFLDKLQKLYPEIFKEVMEASGKTIDFSVKIKWEAKY 1051
QY 1053 FNGTILGRGVYLVKDNADSKYFELKGNQTVLPKQMT-NKEASTGFVNDGNGMTPYSTS 1111
Db 1052 FNGTINIKRQSGSYLVLSL--GKLYFTVNDKGTFLPALTGDTRAKTGFAYDGTGVITYTTS 1109
QY 1112 GYQAANSFVODAKGNWYFPDNGHMYGLQQLNGEVQYFSLNGVQLRESFLENADGSKNY 1171
Db 1110 GTQAQSQFV-TYNGKQYFNDKGYLVTGQTDIGSNYFFLPNGWFTDGRVNRKAGQSULV 1168
QY 1172 FGHNGR-----YSGNY----- 1183
Db 1169 YKSGKLTQTGKWEVTVKDSGKEKFKYQFFKGGIMATGLTEVEGKEKYFYDNGYQAK 1228
QY 1184 -----YS--FNDNSKRWYFDASGVMAVL----- 1205
Db 1229 GVFVPTKDHLMFFCGDSGERKYSGFQEDGNNWYANDKGYVATGTFKVGKQNLVNEKG 1288

QY 1206 -----KTINGNTQYFDQGYOVKGAWITGSDGKKRYFD 1238
Db 1289 VOYKNRFFQVDATYANNEDGVLRGAQTINGDELIFDESGKQVKGFEVANNPDGTTSYD 1348
QY 1239 DSGGNMAVNR-----FANDKNGDWYLLNSDGIALVGVQTIIN 1274
Db 1349 AITGVKLVDTSLTVVDGQTFNVDAKVGVTKAHTPGFYTTGDNNNFYADSYGRNVTTGAQVIN 1408
QY 1275 GKTYIFGQDGKQIKGLIITD-NGKLYFLANSSELARNIIFATDSQNNWYIFGSDGVAVTG 1333
Db 1409 GQHLFYDANGRQVKGFGVTNTDGRSFYHWTGDKLVSTFFATGHDHRYADDRGNVVTG 1468
QY 1334 SQTIAKGLYFASDGQVKGSGFVT-YNGKVHYHADSGLQVNRFEADKQGNWYLLDSNG 1392
Db 1469 AQVINGOKLFFDTDGQVKGAFATNANGRSYHWTNGKLVSTFFTSFGDNNWYADAKG 1528
QY 1393 EALTGSQRINDQVRFFTRREGKQVKGDA 1420
Db 1529 EVVVGQTINGQHLYPDQTKQVKGATA 1556
RESULT 10
T31098
probable dextranucrase (EC 2.4.1.5), extracellular - Leuconostoc mesenteroides
C:Species: Leuconostoc mesenteroides
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T31098
R:Monchole, V.; Renaud-Simeon, M.; Monsan, P.; Willemot, R.M.
FEMS Microbiol. Lett. 159, 307-315, 1998
A:Title: Cloning and sequencing of a gene coding for an extracellular dextranucrase (D)
A:Reference number: Z20981; MUID:98164374; PMID:9503626
A:Accession: T31098
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1508 <MON>
A:Cross-references: UNIPROT:O52224; UNIPARC:UPI00000BB69B; EMBL:AF030129; NID:g2766611;
A:Experimental source: strain NRRL B-1299
C:Genetics:
A:Gene: derB
C:Function:
A:Description: produces dextran composed only of alpha(1-6) glucosidic bonds
C:Keywords: glycosyltransferase; hexosyltransferase
Query Match 43.6%; Score 3252.5; DB 2; Length 1508;
Best Local Similarity 44.3%; Pred. No. 1.5e-134;
Matches 670; Conservative 236; Mismatches 477; Indels 129; Gaps 23;
QY 2 ETKRRYKMKHKVKKHWVTVAVASGLITLGTTLGSSVSAETEQTSKDVVTQKSEDDKAAS 61
Db 6 ERNVRKLYKSGKSWIGLILSTIMLSMTATSONVADSTNTVTDKSVTVSNSTNTNQ 65
QY 62 ESSQTDAPKTKAQTEQTOAQSOAN-----VADTSTSITKETPSQNTTQANSDDK- 112
Db 66 HDTVVD-KQTIPIVKNDQITQQAANATAQAEKVASDITTDQKAEATANNKEDSIDNL 124
QY 113 -----TVNTKSEEAQTSSEERTKQSEEAQTASSQAL- 152
Db 125 TKQLPAVPTANQKTGYLEKQKQKYYVTSNDTLKGLTVDNHNKQYFDNNGVQAKGQFVT 184
QY 153 KQKQK-----AAQENKNPV--DLAAIPN----- 173
Db 185 DNSKTYLLDPNSGNVATGIIQIGSQTLAFDNDNGEQVADFPYFAPDGKTYFDDKGOATIG 244
QY 174 VKQIDGKYYIGSDGQPKKFNALTVNNKVLYPKNTGALTDTSQVQFKOGLTKLNNDYTP 233
Db 245 LKAINGHNYFDSLGLKKGFTGVIDGQVRYDQSGQEVSTDSQIKGLTSTQTTDYTA 304
QY 234 HNQIVNFENTSLETIDNYVTADSWYRPKDILKNGKTTWTASSSDLRPLLMWPDQTOI 293
Db 305 HNAVHSTDSADPNFNGVLTASSWYRPKDVLRNGHWEATTANDFRPIYVVMWPSQTOV 364
QY 294 AYLNMNQGL-GTGENVTADSSQSLNLAATQVQVKIETKISQTOQTQWLRLDIINSFVK 352

QY	1122	DAKGNWYFDNNHGMVYGLQQLNGEVQVYFLSNGVQLRBSFLENADGSK-NYFCHGLNRYVS	1180
Db	1078	EISGSWYFSDSGKMATKTKICNDTYLFPMPNGKQLKEGV--YDGKAYYYDDNGRTWT	1135
QY	1181	N-GYYSF--DNDSKRYFPDASGVMAVGLKTINGTOYFDODGVQVKGAWITSGDGKKRYF	1237
Db	1136	NKGVEFRVDGQDKWRYFNGDGTIAIGLSLDRNLTYFDAYGYQVKGQTVT-INGKSYTF	1194
QY	1238	DDGSGMNAVRFAENDK-----NGDWYLSNSDGIALVGVQTINGKTYFFGQDGKQ	1286
Db	1195	DADQGLVOTDNANPAQOAGWKLGDQWGY-RKGQQLLTGEQITIDGQKVFQDNGVQ	1253
QY	1287	IKGKIITD-NGKLKYLFLANSBELARNIPATDSQNNWYFSGDGVAVTGSQTIAGKKLYPA	1345
Db	1254	VKGGTATDASGLRFPYDRDQGHQVGKWTSTSDNNWYVNESQVLTGLQITIDGQTVYFD	1313
QY	1346	SDGQVKGSGFV-TYNGKVHYHADSGELQVNRFEADKQGNWYFLDSNGEA	1394
Db	1314	DKGIOAKGRAVWDENGLRYFDADSGNMLRDRWK-NVDGNWYFNRNGLA	1362
RESULT 13			
B97033			
uncharacterized protein, related to enterotoxins of other Clostridiales [imported] - Cl			
C:Species: Clostridium acetobutylicum			
C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004			
C:Accession: B97033			
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,			
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.			
J. Bacteriol. 183, 4823-4838, 2001			
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo			
A:Reference number: A96900; MUID:21359325; PMID:21359325			
A:Accession: B97033			
A>Status: preliminary			
A:Molecule type: DNA			
A:Residues: 1-2817 <KUR>			
A:Cross-references: UNIPROT:Q97K42; UNIPARC:UPI00000CA0A0; GB:AE001437; PIDN:AAK79053.1;			
A:Experimental source: Clostridium acetobutylicum ATCC824			
C:Genetic:			
A:Gene: CAC1079			
Query Match 7.5%; Score 562; DB 2; Length 2817;			
Best Local Similarity 21.9%; Pred. No. 2e-16;			
Matches 351; Conservative 206; Mismatches 557; Indels 487; Gaps 77;			
QY	85	ANVADTSTITKTPSON-----ITTOAN-SDDKTNTKSEEAQTSERTKQS	132
Db	26	SELTKNSSALTKRSSNNFSLNKHVPTPITSNVNGSNAKNNLTK-VQNTASSMPTT	84
QY	133	EEAQTTASSQALTQAKAELTKORQTAQENKPNVDLAAIPNVKQIDGKYIYIGSDGQPKK	192
Db	85	NPKQATNSKILVNPKL----NQASSPNEGITPKQASIPYTNVTDKNKTF-----K	132
QY	193	NFALTNNKVLVPDKNTGALTDTDSQYQFKGLTKLANDYTPHNQIVNFNTSLETIDNVY	252
Db	133	NBS-SINNEAPIPKDTSKTSKTSQAQ-----TKGSD-----NNIPS-NNITSTNTSKNEN	181
QY	253	TADSWRPDKLNGKGTWTASSESDLRPLLMSWMPDKQTOIAYLNVNMQGLGTGENYTA	312
Db	182	PSNT-----DI-----KTTEPANAIPKDT-----PNNQSDSA-----LAKNKALSNRNLA	223
QY	313	DSQESLNLAAQTQVVKIETKISQTOQTOWLRIINSFVK-----TQPNWNSQ	360
Db	224	DSSQTSKVITSSNNDAPKVNVTSTDKKASLNNDSDQGWTKDKGKYIYVNGVQKGFSQI	283
QY	361	TSDTSGAGEKDHQGALLYS-----NSDKTAYA-----	389
Db	284	NKSIYFNDGSMQGTWLNKSNSSYFDASGVMLTGLQINGTYFFNDDGKLLTGLQAI	343
QY	390	NSDYRLNRTPTTSQTG----KPKYFEDNS-----SGGYDFLLANDIDNS	429
Db	344	NNYIYFNDGVMTGWITCNDSKYIFDNGVMQGLVHNNKYGFNGDKLLTGLQNI	403

QY	430	NPVQAEQLNWLHYLMNYGSIIVANDPEANFDFGVRVDAVD---NVNADLLQIASDYLKASY	486
Db	404	N-----NTYIYFDSNGVM-----QTDWITIDSKYIYSFVNG-VMTQGIYISGY	447
QY	487	GVDKSEKNAIHLNLSLEAWSNDNDPOYNKDTKGAQLPIDNKLRLSLLYALTRPLEKASNK	546
Db	448	YGFANDGKLLTGLQVI-----NGNSIYFDTNGIRL-----VSRWITIDCKY	489
QY	547	NEIRSLGFBPVIITNSLNRSAECK-----NSERMANYIFIRAHDE--VQTV	590
Db	490	YFNQDG---ILTDNWIN--YDGKYIYISGVKQTKGLQNI--GNYYF--DSSGIMQTG	539
QY	591	IAKIITKAQINPKTDGLTFLDE--LKQAFKI-----YNED--MRQAKKYTQSNIFT	638
Db	540	LOKI-----DGKTYFYGDNIGIRQIGWITTYQNNKYFNSDGSMTDLKYSYSTSY	590
QY	639	AYALMLSNKDSITRLYYGDMYSDDG-----OYMATKSPYYDA--IDTL---LKARIKY	686
Db	591	NTHYQ-----YYG--FNDGKLLTGLQTIKNTYIFDSNGISQMGWNIDGKDFY	638
QY	687	AAGGQDMKITIYVEGDKSHMDWDTGVLTSVRYGTGANEATDQGSSEATKTQGMVITSNNP	746
Db	639	FNSNSIMTENWVINDKYYFYINNKKQTKGFQYKYYFV--PDGIMQTKGFTISGNTY	696
QY	747	SLKLNQNDKVIIVNMG-AAHQNOEYRPLLLTTKDGKTSYTSDDAAKSLYKRTNDKGELVFD	805
Db	697	YL-----DDNGVKQTKGWTIKGKY-----YFDGNGVMINY-----WVFD	731
QY	806	ASDIQGYLNPQV-SGYLAVWVPVGASDNQDVRVAASNKANATQGVYESSALDSOLI-YE	863
Db	732	NDKTYIYINGNMQTKGAISSLNHHYFDDNGIMQTKQWRINGRTYTFDNNGAKTGLVYIE	791
QY	864	G-----FSNFQDFVTKDSYTNKKIAQNVQLFKSWGVSFEMAPQVYS-SEDSGFLDS---	915
Db	792	GKTYIYFNTYAYLDTGFIYFN---NNYYFLDNNGV---VRTGWINYNNRYLDSTGV	843
QY	916	-----IIQNGYAFEDRYDLAMS---KNNKYGSQODMINAVKALHKSGIQVIAWVFD	964
Db	844	RVTGFTIDGNKYIYFDSGAMCTSFITVNGNTYGFSD-----GIM-----	884
QY	965	QIYNLPGEVVTATRVNDYGEYRKDSE-----KNTLYAANTKSGKDYQAKYGG	1014
Db	885	-----ITGWTIILUSSNVSSNYIYFNSDGSQAQGFYTLGKTY-----FEPNYGY	930
QY	1015	AFL--SELAAKYPSIFNRTQISNGKKIDPSEK-----ITAMK-----AKYFNGTNI	1058
Db	931	MLLGYNINGKYIYFNDGVIQTGWVTDRSSKVIYLDPSGAAVTGFQNINGDKYFNSSGI	990
QY	1059	LORGVGVVLKD-----NASDKYF-----LKGNOTYLPK-QMTNKEAS	1095
Db	991	MOTGLVYVNPDIYFGFDDNGHILTMHSHINGIYIYFDSTGKAQKGFYTLGKTYFNTNMY	1050
QY	1096	TCFVNDGNGMTFYSTSG-----YQAKNSFV---QDAKGNWYFDNNHGMV	1137
Db	1051	TCFVNANNMLIYFDNEGVMQGWINVNSRYFSAATGASVTGFTIDGKNCYFDSNGAIY	1110
QY	1138	YGLQQLNGEVQVFLSNGVOLRESFLENADGSKNYFGLHGNRYSNGY-----YSFOND--	1189
Db	1111	TDVVTINGSTYGFNTDGIMLT-----GWQIRYNRGYSSYFNTYFNSDGT	1156
QY	1190	-----SKRYFPDAS-GVMVAGLKTINGNTQYFDQDGVQVKGAWITSGDKKRYFDD	1239
Db	1157	AKTGFTYLNKTYIYFNPSDGRMLQGYIINGNHYYFAPDGTMTQTG-WIT--NGSSKYYLD	1213
QY	1240	GSGNMVAV-----NRFANDKNG-----DWYLSNSDGIALGVQTINGKTYF	1280
Db	1214	PSGAAVTGLQTINGNKYCFDSNGILOHNGIFIGTYNTYGSDDNNGIMLTGLQLINGLYCF	1273
QY	1281	GQDGKQIKGKIITDNGKLYFLANSBELARNIPATDSQNNWYFSGD-----	1327
Db	1274	NSDG-SVKTGLVTLGKTYFDSYSVSGFQNI-----NNNTYFNGDGTMTGWNVGY	1327

Db 1391 DFGSDIDNKDRYIFLTCELDKSLIEINLVAKSYSLLSG-----DKNY 1436
QY 838 AASNKANATGQVYESSALDSQLI---YEGFSNFPDF--VTKDSYV---NKKIAQNVQL 889
Db 1437 LISNLNTIEKI--NTLGLDSKNIAINYTDENSNKYFGAISKTSQKSIIHYKDSKNILE 1494
QY 890 FKSGVTSPEMARQYVSSDSGL-DSI--IQNGYAFEDRYDLA-----MSKN----- 934
Db 1495 FYNDSTLEFN-SKDFIADINVFMDKDIINTITGKYVDVNNNTDKSIDFSISLVSKNQKVN 1553
QY 935 -----NKYGSQODMINAVKALHSG-----IQVIADW-----VPODIYNLPGK 972
Db 1554 GLYLNESVYSSYLDLFFVNSDGHNTSNFMNLFDLNISFWKLFCFENINVIDKYFTLVGK 1613
QY 973 -----EVTATRVN---DYGEYRKDSEIKNTLYAANTKS-----NGKD----- 1007
Db 1614 TNLGYVEFICDNNKNDIYFGEW-KTSSSKSTIFSGNGRVVVEPIVNPDTGEDISTSLD 1672
QY 1008 --YQAKYG-----GAFSELAAKYPSIPNRTQISNGKK 1038
Db 1673 FSYEPLYGIDRYINKVLIAPDLTSLININTNYSNEYYPEIIVLNPNTFHKKVWIN--- 1729
QY 1039 IDPSEKITAWK-----AKYFNGTN-----ILGRG 1062
Db 1730 LDSSEFYKWSGTEGDFILVRYLEESNKKILQIRIKGILSNQTSFNKMSIDFKDIKLS 1789
QY 1063 VGYVL-----KDNASDK---YFELKGNQTYLPKQMTNKEASTGFVNDGNGMTFYS--- 1109
Db 1790 LGYIMSNFKSFENSELDRDLGFKIIDNKTYVYDE--DSKLVKGLININNSLFYDPDIE 1847
QY 1110 ---TSGYQAKNSFVQDAKGNWYFDP--NNGHMVGLQQLNGEVOYFLSNGVOLRESFLENA 1165
Db 1848 FNLVTGQMTIN-----GKKYFEDINTGAALTSYKIIINGKHFFYFNNDGV-MQLGVFKGP 1899
QY 1166 DGSKNYFGLHGNYSN-----GYYSFENDSK-----WR-----YEDA 1197
Db 1900 DGFE-YFAPANTQNNIEGQAIYVQSKFLTNGKKYTFDNNSKAVTGWRILINNEKYFNP 1958
QY 1198 -SGVMAVGLKTINGNTQYFDQDGQYQVKGAMITGSDGKGRYFDGSGNMVAVNRPANDKNGD 1256
Db 1959 NNAIAAVGLQVIDNNKYNFNPDTAIIISKGWQT-VNGSRYFYDDTDA-IAPNGY-KTIDGK 2015
QY 1257 WYILNSDGIALGV-----QTINGKTYFFQDGK 1285
Db 2016 HFYFDSQVVKIGVFSTSGFEYFAPANTYNNIEGQAIYVQSKFLTNGKKYTFDNNSK 2075
QY 1286 QIKGKIITDNGKLKYFLANSSELARNIFATDSQNNWYFGSD-GVAVTGSOTIAGKKLYF 1344
Db 2076 AVTGWQTIDSKKY-YFNTNTAEATGWQTIDGKK--YFNTNTAEATGWQTIDGKKYF 2132
QY 1345 ASDGKQVKSFTVYNGKVVHYHADSGELQV-----NRFE-----ADKD----- 1382
Db 2133 NNTWTAISTGYTTIINGK-HFYFNTDGIQIGVFKPGNPFYFAPANTDANNIEGQAILYQ 2191
QY 1383 -----GNWYILDSNGEALTGSQRINDQRVFF 1408
Db 2192 NEFLTNGKKYFSGDSKAVTGWRIINNKYYF 2224

RESULT 15
A37184
glucan-binding protein - Streptococcus mutans
C;Species: Streptococcus mutans
C;Date: 31-Jan-1992 #sequence_revision 31-Jan-1992 #text_change 09-Jul-2004
C;Accession: A37184
R;Banas, J.A.; Russell, R.R.B.; Ferretti, J.J.
Infect. Immun. 58, 667-673, 1990
A;Title: Sequence analysis of the gene for the glucan-binding protein of Streptococcus m
A;Reference number: A37184; PMID:90170123; PMID:2307516
A;Accession: A37184
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-563 <BAN>

A;Cross-references: UNIPROT:Q54447; UNIPARC:UPI000008BCB5; GB:M30945; NID:g153637; PIDN
F;169-188/Domain: cpl repeat homology <CP1>
F;264-283/Domain: cpl repeat homology <CP2>
F;349-368/Domain: cpl repeat homology <CP3>
F;504-523/Domain: cpl repeat homology <CP4>
F;525-548/Domain: cpl repeat homology <CP5>
Query Match 5.8%; Score 435, DB 2; Length 563;
Best Local Similarity 26.8%; Pred. No. 6.8e-12;
Matches 137; Conservative 72; Mismatches 180; Indels 122; Gaps 19;
QY 989 DSEIK-NTLYAANTKSNKGKDYQAKYGAFSELAAKYPSIENRTQISNGKKIDPSKITA 1047
Db 70 DSVENNSANAVKSBETAEEVSDGR-----ASQTEAVTNGT-----NSEEHHHPAEKATA 121
QY 1048 WKAKYFNGTNIILGRGVGYVLKDNASDKYFELKGNQTYLPKQMTNKEASTGFVNDGNGMTF 1107
Db 122 VSGE-----AQSVQNAPESEAAQ-----QETAKTEPATAAENNDAAPT- 159
QY 1108 YSTSGVQAKNSFVQDAKGNWYFDPNNGHMVGLQQLNGEVOYFLSNGVOLR-ESFLENAD 1166
Db 160 -----NSPFK-KDGKWKYKADGQLATGQIIDGRKQLYFNQDGSQVKGIEIHVETGD 209
QY 1167 GSKNYFGLHGNYS---NGYYSFENDS-----KWEYFEDA 1197
Db 210 QIIYHPVFISSPSVLEVNKIYFDPDSELKWDKRFVYSSYADPLHYENIKHEGWLYLGE 269
QY 1198 SGVMAVGLKTINGNTQYFDQDGQYQVKGAMITGSD-----GKCRYFDGSGNMVAVNR 1248
Db 270 DGKAAIGWRTIGKKYFDTNGVQVKGKLISTDGNYNLISQKYGKKSFLDPDTGEAWTNR 329
QY 1249 FANDK-----NGDWYILNSDGIALGVQTINGKTYF-GQDGKQIKGIITDNG 1296
Db 330 FVNAKTYFYNFAGYVSTTDFWYMGADGIGVTDWQKIDGMDYFPEPSSGIVQKGDIAERDG 389
QY 1297 KLKYFLANSSELARNIFATDSQ-----NWYFGSDGVA---VTGSOTIAGKKLYF 1344
Db 390 KVIYLEDSDSGQVVKRFRFGTTPAERISTVEARFPKTYFAGDSRKDLTGWQIIDGKTYF 449
QY 1345 ASDGK-QVKGSFVYNGKV-----HYHADSGELQVNR--EADKGNWYILDS 1390
Db 450 KDDHSIKAKSEYSQIGSVDPDGPFAIDGDGYFFDQGGQVFNTRFVRKYDYSNIWYIYGS 509
QY 1391 NGEALTGSQRINDQRVFFTRE-----GKQVKG 1417
Db 510 DGRVSGWQTIDGKRYFSDQEKTKGRQIKG 540

Search completed: February 11, 2006, 19:42:03
Job time : 38.5158 sec

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 18:57:35 ; Search time 192.956 Seconds
(without alignments)
5228.676 Million cell updates/sec

Title: US-10-797-821-36
Perfect score: 7462
Sequence: 1 METKRYKMKVKKHWTVA.....EGKQKGVAYDERRLLVYR 1430

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Uniprot 05.80.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7412	99.3	1462	1 GTFD_STRMU	P49331 streptococc
2	4550.5	61.0	1577	2 Q54178 STRGN	Q54178 streptococc
3	4445.5	59.6	1575	2 Q9LCH3 STOR	Q9LCH3 streptococc
4	3860	51.7	1476	1 GTFB_STRMU	P08987 streptococc
5	3841.5	51.5	1454	2 Q69A94 LEUME	Q69A94 leuconostoc
6	3788.5	50.8	1577	2 Q55265 STRSL	Q55265 streptococc
7	3743	50.2	1455	1 GTFC_STRMU	P13470 streptococc
8	3713.5	49.8	1449	2 Q68542 STRSL	Q68542 streptococc
9	3713.5	49.8	1449	2 Q55264 STRSL	Q55264 streptococc
10	3661	49.1	1512	2 Q9WXJ5 STRE	Q9WXJ5 streptococc
11	3639	48.8	1506	2 Q56CX8 STRE	Q56CX8 streptococc
12	3615	48.4	1554	2 Q8KZL5 STRE	Q8KZL5 streptococc
13	3587.5	48.1	1597	1 GTF1_STRDO	P11001 streptococc
14	3575	47.9	1590	2 Q59983 STRE	Q59983 streptococc
15	3544	47.5	1590	2 Q55263 STRE	Q55263 streptococc
16	3542	47.5	1592	1 GTF2_STRDO	P27470 streptococc
17	3502.5	46.9	1518	2 Q00600 STRSL	Q00600 streptococc
18	3369.5	45.2	1561	2 Q55BM8 LACO	Q55BM8 lactobacill
19	3334.5	44.7	2835	2 Q8G9Q2 LEUME	Q8G9Q2 leuconostoc
20	3312	44.4	1599	2 Q00599 STRSL	Q00599 streptococc
21	3252.5	43.6	1508	2 Q52224 LEUME	Q52224 leuconostoc
22	3251.5	43.6	1290	2 Q48756 LEUME	Q48756 leuconostoc
23	3245.5	43.5	1508	2 Q9ELH5 LEUME	Q9ELH5 leuconostoc
24	3227	43.2	1477	2 Q91466 LEUME	Q91466 leuconostoc
25	3186.5	42.7	1527	2 Q8KRE1 LEUME	Q8KRE1 leuconostoc
26	3174	42.5	1330	2 Q84CN4 LEUME	Q84CN4 leuconostoc
27	3174	42.5	1365	1 GTFS_STRDO	P29336 streptococc
28	3172.5	42.5	1522	2 Q6TXV4 LEUME	Q6TXV4 leuconostoc
29	3156.5	42.3	1527	2 Q9XAR4 LEUME	Q9XAR4 leuconostoc
30	3105.5	41.6	1338	2 Q9WXJ4 STRE	Q9WXJ4 streptococc
31	3067	41.1	1595	2 Q55BM3 LACSK	Q55BM3 lactobacill

RESULT 1

GTFD_STRMU

ID GTFD_STRMU STANDARD; PRT; 1462 AA.

AC P49331; O69383; O69386; O69389; O69392; O69398;

DT 01-FEB-1996 (Rel. 33, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 10-MAY-2005 (Rel. 47, Last annotation update)

DE Glucosyltransferase-S precursor (EC 2.4.1.5) (GTF-S) (Dextranucrase)

DE (Sucrose 6-glucosyltransferase).

OS Name=gtfd; OrderedLocustNames=SMU.910;

GN Streptococcus mutans.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI_TaxID=1309;

OX [1]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=GS-5;

RC MEDLINE=91100958; PubMed=2148600;

RA Honda O., Kato C., Kuramitsu H.K.;

RA "Nucleotide sequence of the Streptococcus mutans gtfd gene encoding the glucosyltransferase-S enzyme";

RT J. Gen. Microbiol. 136:2099-2105(1990).

RL [2]

RN NUCLEOTIDE SEQUENCE.

RP STRAIN=MT4239 / Serotype c, MT4245 / Serotype e, MT4251 / Serotype f, MT4467 / Serotype e, and MT4148 / Serotype c;

RC MEDLINE=98231643; PubMed=9570124;

RX Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S., Kimura S., Hamada S.;

RA "Molecular analyses of glucosyltransferase genes among strains of Streptococcus mutans.";

RL FEMS Microbiol. Lett. 161:331-336(1998).

RN [3]

RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

RC STRAIN=UA159 / ATCC 700610 / Serotype c;

RA MEDLINE=2295063; PubMed=12397186; DOI=10.1073/pnas.172501299;

RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J., Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P., Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A., Ferretti J.J.;

RA "Genome sequence of Streptococcus mutans UA159, a cariogenic dental pathogen.";

RT Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).

RL CC -!- FUNCTION: Production of extracellular glucans, that are thought to play a key role in the development of the dental plaque because of their ability to adhere to smooth surfaces and mediate the aggregation of bacterial cells and food debris.

CC CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl)(n) = D-fructose + (1,6-alpha-D-glucosyl)(n+1).

CC CC -!- SUBCELLULAR LOCATION: Secreted.

CC CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both forms of glucans.

32 2723.5 36.5 1463 2 Q55BM6 LACFE

33 2699.5 36.2 2057 2 Q9RE05 LEUME

34 2642 35.4 1016 2 Q9LCJ7 LEUME

35 2460.5 33.0 1772 2 Q55BN3 LACRE

36 2447 32.8 1781 2 Q55BL9 LACRE

37 2408.5 32.3 1781 2 Q4JCS4 LACRE

38 2405 32.2 1772 2 Q55BN0 LACRE

39 2404.5 32.2 1781 2 Q4JLC7 LACRE

40 1552.5 20.8 591 2 Q8VUH3 STRMU

41 1344.5 18.0 1619 2 Q55BM0 LACRE

42 1330.5 17.8 1231 2 Q55BN1 LACRE

43 1205 16.1 522 2 Q8VV10 STRSA

44 823 11.0 374 2 Q6ZX19 LACO

45 562 7.5 2817 2 Q97K42 CLOAB

Q55bm6 lactobacill

Q9re05 leuconostoc

Q9lcj7 leuconostoc

Q55bn3 lactobacill

Q55bl9 lactobacill

Q4jcs4 lactobacill

Q55bn0 lactobacill

Q4jlc7 lactobacill

Q8vuh3 streptococ

Q55bm0 lactobacill

Q55bn1 lactobacill

Q8vv10 streptococ

Q6zx19 lactobacill

Q97k42 clostridium

CC CC -|- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
 CC CC -|- SIMILARITY: Contains 6 cell wall binding repeats.
 CC CC -----
 CC CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC CC the European Bioinformatics Institute. There are no restrictions on its
 CC CC use as long as its content is in no way modified and this statement is not
 CC CC removed.
 CC CC -----
 CC EMBL; M29296; AAA26895.1; -; Genomic DNA.
 DR DR EMBL; D88653; BAA26103.1; -; Genomic DNA.
 DR DR EMBL; D88656; BAA26107.1; -; Genomic DNA.
 DR DR EMBL; D88659; BAA26111.1; -; Genomic DNA.
 DR DR EMBL; D88662; BAA26115.1; -; Genomic DNA.
 DR DR EMBL; D89979; BAA26121.1; -; Genomic DNA.
 DR DR EMBL; A8014932; AAN58619.1; -; Genomic DNA.
 DR DR HSSP; P06653; LGVM.
 DR DR InterPro; IPR002479; Cell wall bd put.
 DR DR Pfam; PF003318; Glyco_hydro_70.
 DR DR Pfam; PF01473; CW_binding_1; 4.
 DR DR Pfam; PF02324; Glyco_hydro_70; 1.
 KW Complete proteome; Dental Caries; Glycosyltransferase; Repeat; Signal;
 KW Transferase.
 FT FT SIGNAL ? Potential.
 FT FT CHAIN ? Glucosyltransferase-S.
 FT FT REPEAT 1232 1295 1.
 FT FT REPEAT 1296 1359 2.
 FT FT REPEAT 1360 1423 3.
 FT FT REGION 1232 1423 3 X 63 AA approximate tandem repeats.
 FT FT VARIANT 10 10 Y -> H (in strain GS-5, strain MT4239,
 FT FT strain MT4245, strain MT4251, strain
 FT MT4467 and strain WT8148).
 FT VARIANT 19 19 I -> V (in strain GS-5, strain MT4239,
 FT strain MT4245, strain MT4251, strain
 FT MT4467 and strain WT8148).
 FT VARIANT 58 58 K -> E (in strain MT4467).
 FT VARIANT 68 68 A -> S (in strain MT4239 and strain
 FT MT4245).
 FT VARIANT 81 81 A -> T (in strain MT4251 and strain
 FT MT8148).
 FT VARIANT 113 113 T -> I (in strain MT4239 and strain
 FT MT4245).
 FT VARIANT 122 122 A -> V (in strain MT4239, strain MT4245
 FT and strain WT8148).
 FT VARIANT 132 132 A -> S (in strain GS-5 and strain
 FT MT4467).
 FT VARIANT 135 135 A -> V (in strain MT4245).
 FT VARIANT 137 137 A -> T (in strain GS-5, strain MT4239,
 FT strain MT4245, strain MT4251, strain
 FT MT4467 and strain WT8148).
 FT VARIANT 202 202 V -> L (in strain MT4239).
 FT VARIANT 255 255 D -> N (in strain WT8148).
 FT VARIANT 275 275 E -> D (in strain MT4239, strain MT4245
 FT and strain MT4251).
 FT VARIANT 288 288 D -> N (in strain MT4239, strain MT4245
 FT and strain MT4251).
 FT VARIANT 301 301 Q -> H (in strain MT4245).
 FT VARIANT 313 313 D -> N (in strain MT4239 and strain
 FT MT4251).
 FT VARIANT 317 317 E -> K (in strain MT4239).
 FT VARIANT 328 328 V -> F (in strain MT4239).
 FT VARIANT 350 350 F -> L (in strain MT4239, strain MT4251
 FT and strain MT4467).
 FT VARIANT 628 633 KKRYTQ -> EKEYTL (in strain MT4251).
 FT VARIANT 688 688 A -> S (in strain MT4239).
 FT VARIANT 726 732 TDGSEA -> ADKGND (in strain MT4251).
 FT VARIANT 726 730 TDGGS -> ADKGN (in strain MT4239 and
 FT strain MT4245).
 FT VARIANT 762 762 T -> A (in strain GS-5, strain MT4239,
 FT strain MT4245, strain MT4251, strain
 FT MT4467 and strain WT8148).
 FT VARIANT 964 964 D -> Y (in strain MT4251).
 FT VARIANT 1019 1019 E -> K (in strain MT4245 and strain

MT4251).
 LG -> IR (in strain MT4251).
 G -> R (in strain MT4245).
 H -> Q (in strain MT4239).
 S -> N (in strain GS-5).
 Y -> C (in strain MT4239).
 Y -> C (in strain MT4251 and strain
 MT4467).
 F -> L (in strain MT4467).
 Q -> P (in strain MT4245).
 K -> T (in strain MT4245).
 N -> D (in strain MT4245).
 G -> D (in strain GS-5 and strain
 MT4467).
 G -> R (in strain GS-5).
 R -> K (in strain MT4467).
 RYDKNSGMVYKVVVTLANGRRIGIDRWGIARY -> VY
 R (in Ref. 1).
 SQ SEQUENCE 1462 AA; 163388 MW; CE4A279C4D708645 CRC64;
 Query Match 99.3%; Score 7412; DB 1; Length 1462;
 Best Local Similarity 99.4%; Pred. No. 5.4e-304;
 Matches 1420; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
 QY 1 METKRYKMHKVKHVVAVASGLITLGTTLGSSVSAETEQOTSDKVVYTKSEDDKAA 60
 DB 1 METKRYKMYKVKHVVAVASGLITLGTTLGSSVSAETEQOTSDKVVYTKSEDDKAA 60
 QY 61 SESSQTDAPKTQAQTEQTAQSQANVADTSTSITKETPSQNTTQANSDDKTVTNTKSE 120
 DB 61 SESSQTDAPKTQAQTEQTAQSQANVADTSTSITKETPSQNTTQANSDDKTVTNTKSE 120
 QY 121 EAQTSERTKQSEEAQTASSQALTAQAKELTKQRTAAQENKPNVDLAAIPNVKQIDGK 180
 DB 121 EAQTSERTKQSEEAQTASSQALTAQAKELTKQRTAAQENKPNVDLAAIPNVKQIDGK 180
 QY 181 YYYISGDGPKKNFALTNNKVLVFDKNTGALTDTTSQYQKGLTKLNNDYTPHNOIVNF 240
 DB 181 YYYISGDGPKKNFALTNNKVLVFDKNTGALTDTTSQYQKGLTKLNNDYTPHNOIVNF 240
 QY 241 ENTSLETIDNVYTADSWYRPKDIKNGKTTWTASSESDLRPLLMWMPDKQTOIAYLNYMN 300
 DB 241 ENTSLETIDNVYTADSWYRPKDIKNGKTTWTASSESDLRPLLMWMPDKQTOIAYLNYMN 300
 QY 301 QQGLGTGENYTADSSQESLNLAQTQVVKIETKISQTOQTQWLRLDIINSFVKTPQPNWSQ 360
 DB 301 QQGLGTGENYTADSSQESLNLAQTQVVKIETKISQTOQTQWLRLDIINSFVKTPQPNWSQ 360
 QY 361 TESDTASAGEKHLQGGALLYSNSDKTAYANSYRLLNRTPTTSQTKPKYFEDNSSGGYDF 420
 DB 361 TESDTASAGEKHLQGGALLYSNSDKTAYANSYRLLNRTPTTSQTKPKYFEDNSSGGYDF 420
 QY 421 LLANDIDNSNPVQAEQLNWLHLYLMNYGSIIVANDPEANPDGVRVDAVDNVDNLLQIASD 480
 DB 421 LLANDIDNSNPVQAEQLNWLHLYLMNYGSIIVANDPEANPDGVRVDAVDNVDNLLQIASD 480
 QY 481 YLKAHYGVDKSEKNAIHLISLEASDNDPQYNKDTKGALPIDNKLRLSLLYALTRPLE 540
 DB 481 YLKAHYGVDKSEKNAIHLISLEASDNDPQYNKDTKGALPIDNKLRLSLLYALTRPLE 540
 QY 541 KDAANKNEIRSGLEPVITNSLNNSAEGKNSERMANYIFIRAHDSVQTVIAKIIKAQIN 600
 DB 541 KDAANKNEIRSGLEPVITNSLNNSAEGKNSERMANYIFIRAHDSVQTVIAKIIKAQIN 600
 QY 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYYTQSNIPTAYALMLSNKDSITRLYYGDMYS 660
 DB 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYYTQSNIPTAYALMLSNKDSITRLYYGDMYS 660
 QY 661 DQGQYMATKSPYDAIDTLLKARIKYAAGQDMKITTYVEGDKSHMDWDYTGVLTSVRYGT 720
 DB 661 DQGQYMATKSPYDAIDTLLKARIKYAAGQDMKITTYVEGDKSHMDWDYTGVLTSVRYGT 720
 QY 721 GANEATDQGSSEATKTQGMVITSNPPLSKLNQNDKVIIVNMGAAHKNQEVRLPILLLTTKQGL 780

```
Db 721 GANEATDQGSSEATKTOGMVITSNPSLKLQNDKIVVNNGTAKHQEYRPLLLTTKDLG 780
QY 781 TSYTSDAAAKSLYRKTNDKGELVFDASDIQGYLNPQVSGYLAWVPVPGASDNODVRAAS 840
Db 781 TSYTSDAAAKSLYRKTNDKGELVFDASDIQGYLNPQVSGYLAWVPVPGASDNODVRAAS 840
QY 841 NKANATGQVYESSALDSQLIYEGFSNFQDFTKSDYTNKKIAQNVLFPKSGVTSFEM 900
Db 841 NKANATGQVYESSALDSQLIYEGFSNFQDFTKSDYTNKKIAQNVLFPKSGVTSFEM 900
QY 901 APOVYSEDSGLFSDSIIONGYAEDRYDLAMSNNKYGSOODMINAVKALHKSIGIOVIAD 960
Db 901 APOVYSEDSGLFSDSIIONGYAEDRYDLAMSNNKYGSOODMINAVKALHKSIGIOVIAD 960
QY 961 WVPDQIYNLPKGEVWVATRVNDYGEYRKDSEIKNTLYAANTKSNKGDKYQAKYGCAFSEL 1020
Db 961 WVPDQIYNLPKGEVWVATRVNDYGEYRKDSEIKNTLYAANTKSNKGDKYQAKYGCAFSEL 1020
QY 1021 AAKYPSIFNRTOISNGKKIDPSEKITAWEKAKYFNGTNIILGRGVYVLDKNASDKYFELKG 1080
Db 1021 AAKYPSIFNRTOISNGKKIDPSEKITAWEKAKYFNGTNIILGRGVYVLDKNASDKYFELKG 1080
QY 1081 NOTYLPKQMTNKEASTGFVNDGNGMTFYSTGYQAKNSFVQDAKGHWYFDDNCHWYGL 1140
Db 1081 NOTYLPKQMTNKEASTGFVNDGNGMTFYSTGYQAKNSFVQDAKGHWYFDDNCHWYGL 1140
QY 1141 QOLNGEVQVFLSNGVOLRESFLENADGSKNYFCHLGNRSNGYYSFDNDSKNRYPDASGV 1200
Db 1141 QOLNGEVQVFLSNGVOLRESFLENADGSKNYFCHLGNRSNGYYSFDNDSKNRYPDASGV 1200
QY 1201 MAVGLKTINGNTQYFPDQGVQVKGAWITGSDGKKRYFDDGSGNMVNRFANDKNGDWYYL 1260
Db 1201 MAVGLKTINGNTQYFPDQGVQVKGAWITGSDGKKRYFDDGSGNMVNRFANDKNGDWYYL 1260
QY 1261 NSDGIALVGQTINGKTYFYEGDQKQIKGKIITDNGKLYFLANSSELARNIFATDSQNN 1320
Db 1261 NSDGIALVGQTINGKTYFYEGDQKQIKGKIITDNGKLYFLANSSELARNIFATDSQNN 1320
QY 1321 WYFSGDGVAVTGSQTIAGKCLYFASDGQVKGSGFVYTKGKHVHYHADSGELQVNRFEAD 1380
Db 1321 WYFSGDGVAVTGSQTIAGKCLYFASDGQVKGSGFVYTKGKHVHYHADSGELQVNRFEAD 1380
QY 1381 KQGNWYILDSNGEALTGSRINDQRVFTFREGKQVKGVDVAYDERLLVY 1429
Db 1381 KQGNWYILDSNGEALTGSRINDQRVFTFREGKQVKGVDVAYDERGLLY 1429

RESULT 2
Q54178 STRGN
ID Q54178 STRGN PRELIMINARY; PRT; 1577 AA.
AC Q54178; Q54247;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Glucosyltransferase.
GN Name=gtfG;
OS Streptococcus gordonii.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1302;
RN [1]
RC NUCLEOTIDE SEQUENCE.
RX MEDLINE=96157064; PubMed=8586185;
RA Vickerman M.M.; Sulavik M.C.; Clewell D.B.;
RT "Molecular analysis of Streptococcus gordonii glucosyltransferase
phase variants."
RL Dev. Biol. Stand. 85:309-314(1995).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Challis;
RA Minick P., Vickerman M.;
```

```
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; U12643; AAC343483.1; -, Genomic DNA.
DR PIR; B41898; B41898.
DR HSP; P06653; 1HCX.
DR GO; CO:0016740; P:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_I; 5_70.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1577 AA; 177805 MW; 5AE0328DC5E08D18 CRC64;

Query Match 61.0%; Score 4550.5; DB 2; Length 1577;
Best Local Similarity 59.7%; Pred. No. 2.5e+183;
Matches 895; Conservative 178; Mismatches 314; Indels 111; Gaps 17;

QY 1 METRRYKHKVKKHVVAVASGLITLGTTLG-----SSV-SA 39
Db 2 MEKKVHYKHKVKKHVVAVTSLALLVAPKALGSLGVYADDAQVTVNVKQSAVQSK 61
QY 40 ETEQOTSDKVVTKSEDDKAASESSQTDAPKTKQAQTEQTOAQSOAN-VADTSTS----- 93
Db 62 DSEQTTSDK-----ATDSSQLKVBQASSSKETYQASAATNPANEQTTOODKE 110
QY 94 -ITKETPSONITQANSDDKTVTNTKSEEAQTSSEETKQSEEAQTASSOALTOAK---- 148
Db 111 VETSRDSRHELTQKTSDD-----SSEKSGSSQEPKQVADQAESTDTKTQAALQAKQDSR 163
QY 149 -----AELTKQRTQATAQENKPV-----DLA 169
Db 164 ANDBETTENVAKATVSDKIATPKKERLPEPAQRKESITEKMLAAQAQAAPVNTHEHDD 223
QY 170 AIPNVKQIDGKYIYGSDGQPKKNFALTNNKVLKFDKVTGALTDFSOYQFKGLTKLNN 229
Db 224 VLAHKITDGKYYVQDDGTVKKNFAVELNGKILYFPAETGALVDSNEYQFQOGTSSLNN 283
QY 230 DYTPHNQIVNFENTSIETIDNTVTADSWYRPKIDILKNGKWTASSSDLRPLLMWMPDK 289
Db 284 EFTQKNAFVGTDKDIEIVDGLVTADSWYRPFILKDGKTWTASTETDLRPLLMWMPDK 343
QY 290 QTOIAYLNYMNOQGLGTGENYTADSSQESLNLAQTVQVKIETKISQTOQTQWLRIINS 349
Db 344 RTQINLYNMNENLIGGA-FESKTEQVLLTNAVQVQRKIEERI SKEGDTKWLRLMSA 402
QY 350 FVKTPQNNNSQTESDTSAGEKHLOGGALLYNSDKTAYANGDYRLINLTPTTSQTKPKY 409
Db 403 FVKTPQNNNIKTESBTTGTGKDHLOGGALLYTNSDKTSHANSRYRLNLTPTTSQTKPKY 462
QY 410 FEDNSSGGYDFLLANDINSNPVQAEQLNLWLHLYMNYGSI VANDPEANFDGVRVDAVDN 469
Db 463 FIDKSNNGGYEFLLANDFNSNPVQAEQLNLWLHLYMNYGSI VANDPTANFDGVRVDAVDN 522
QY 470 VNADLLQIASDYLKAHYGVDKSEKNAIHLSTILEAWSNDNDPQYNKDTKGAQLPIDNKJRL 529
Db 523 VNADLLQIASDYFKSRKYGVSEEEALKHLSILEAWSNDNDPQYNKDTKGAQLPIDNKJRL 582
QY 530 SLLIYALTRPLEKXDAKNKEIRSGLEPVTNSLNNSAEGKNSERMANIYIFIRAHSEVQT 589
Db 583 SLLYSFMRKL-----SIRSGVEPTITNSLNDSTENKNGERTANYIFVRAHDSVQQT 634
QY 590 VIAKIIAQINPKTDGLTFLDELQKAFKIYNEDNRQAKKYTQSNIPAYALMLSNKDS 649
Db 635 VIADIIRENINPNTDGLTFLDELQKAFKIYNEDNRKADKKYQFNIPAHALMLSNKDS 694
QY 650 ITRLYYGDMSYDQGYMATKSPYYDAIDTLKARIKYAAGQDMKITVY----EGDKSHM 705
Db 695 ITRVYVGLYTDGQYMEKSPYDAIDALLRARIKYVAGQDMKVTVYGVPREADK--- 751
QY 706 DWDYTGVLTSVRYGTGANEATDQSEATKQGMVITSNPSLKLQNDKVIWNMAAHK 765
Db 752 -WSYNGILTSVRYGTGANEATDEGTAETRTQGMVITSNPSLKLQNDKVIWNMAAHK 810
```

Qy	766	NQEVRLPILLTTKQGLTGY - TSDAAKASLYKRTNDKGBLVFPDASDIQGYLNPQVSGYLAVW	824
Db	811	NQYVRPVLTTTKDGISRYLTDEEVPQSLWKTKTDANGILTFDMDMDIAGYSNVQVSGYLAVW	870
Qy	825	VPVCSADNPQVRVAASKANATGOVYESSSALDSOLLYEGFSFQDPVTKSDYTNKKIA	884
Db	871	VPVGAKENQDARVVTASKKKNAAGQGVYESSFALDSOLLYEGFSNFQDPFATRDDQVTKNKVIA	930
Qy	885	QNVOLFSGWGTSPFEMAPQVYSSBDGSLDLSIIQNGYAFEDRDYDLAMSKNKTKGSOQDMI	944
Db	931	KNNVLFXEWGVTSELPQVYSSQDGTFLDSIIQNGYAFEDRDYDMAMSKNNKTKGSLDLDL	990
Qy	945	NAVKAHLKSGIQVIADWPVQIYINLPKGEVVTATRVNDYGEYRKDSIKNTLYAANTKSN	1004
Db	991	NALRALHSVNIQAIADWPVQIYINLPKGEVVTATRVNNYGTYREGAEIKEKLYVANSKTN	1050
Qy	1005	GKDYOAKYGGAFLELAAYPSIFNRQIISNGKKIDPSEKITAWEKAYFNGTWTILRGVVG	1064
Db	1051	GTDYQGYGGAFULDELKAKYPEIPERVQISNGQKMTTDEKITKWSAKYFNGTWTILRGAY	1110
Qy	1065	YVLKDNASDKYFELKGNQTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAK	1124
Db	1111	YVLKDWGSKYVLSNKGNETALPKQLVNKEASTGFVKDTNFKFYSTSGNQAKTFFQDEN	1170
Qy	1125	GNWYTFDNNGHMYGLOQLNGEVQYFSLNSGVQLRESFLENADGSKNYFHLGNRYSNYY	1184
Db	1171	GNWYTFDNGYLVGTAREIDGKQLYFMKNGVQLRDALQEDENGNOYYDYDKTGAKVLNRY	1230
Qy	1185	SFNDNSKWRYPDASGVMAVGLKTINGTQYFDDQGYQVKGAWITGSDGKKRYFDDGSGNM	1244
Db	1231	TSDGQN - WRYPDAKGVMAVGLVKIDGQQYFDQNGYQVKGKVRADKGLRYFDDGSGNA	1289
Qy	1245	AVNRFA - NDKNGDWYILNSDGIALVGVQTTNGKTYFYFGDGKQIKGKIITDNGK - LKYFL	1302
Db	1290	VINRFAQDNPSPDWYFYGADGVALTQKTKQOQTYLYFGDGKQVKGQVVMADKSKIRYFD	1349
Qy	1303	ANSGLARNITPATDSQNNWYFGSDGVAVTGQSOTIAGKGLYFASDGGKQVKGFSVITYNGK -	1361
Db	1350	ANSSEMVAKNFAEGAKNEWYFDQDGKAVTGLTKTINNQLYFDDGDKQVKGQVVTIADKS	1409
Qy	1362	VHYTHADSGELQVNRFEADKGNWYILDSNGEALTGSQLRINDQRYFPTREGKQVKGVD	1419
Db	1410	IRYFDANSSEMVAKNFAEGAKNEWYFDQDGKAVTGLQTKINKQVLYFGDGKQVKGQV	1467
RESULT 3			
Q9LCH3_STROR			
ID	Q9LCH3	STROR	PRELIMINARY; PRT; 1575 AA.
AC	Q9LCH3;		
DT	01-OCT-2000	(TrEMBLrel. 15, Created)	
DT	01-OCT-2000	(TrEMBLrel. 15, Last sequence update)	
DT	01-JUN-2003	(TrEMBLrel. 24, Last annotation update)	
DE	Glucosyltransferase.		
GN	Name=gtfr;		
OS	Streptococcus oralis.		
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;		
OC	Streptococcus		
OX	NCBI_TaxID=1303;		
RN	[1]		
RC	NUCLEOTIDE SEQUENCE.		
RC	STRAIN=ATCC10557;		
RX	MEDLINE=20231779; PubMed=10768934;		
RX	DOI=10.1128/TAI.68.5.2475-2483.2000;		
RA	Fujiwara T., Hoshino T., Ooshima T., Sobue S., Hamada S.;		
RT	"Purification, characterization, and molecular analysis of the gene		
RT	encoding glucosyltransferase from Streptococcus oralis.";		
RL	Infect. Immun. 68:2475-2483 (2000).		
DR	EMBL; AB025228; BAA95201.1; -; Genomic_DNA.		
DR	HSSP; P06653; 1HCX.		
DR	GO; GO:0016740; F:transferase activity; IEA.		
DR	GO; GO:0009230; P:glucan biosynthesis; IEA.		
DR	InterPro; IPR002479; CW binding.		
DR	InterPro; IPR003318; Glyco hydro. 70.		

RC STRAIN=IBT-PQ;
RA Fernandez-Vazquez J.L., Lopez-Munguia A., Olvera C.;
RT "Molecular characterization of a dextranucrase gene from Leuconostoc
RL mesenteroides IBT-PQ isolated from pulque.";
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY504865; AAS79426.1; -; Genomic DNA.
DR GO; GO:0047849; F:dextranucrase activity; IEA.
DR GO; GO:0016757; F:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco hydro 70.
DR Pfam; PF01473; CW binding 1; 1_70.
DR Pfam; PF02324; Glyco hydro 70; 1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 1454 AA; 160569 MW; 2DC7342963B6FD87 CRC64;

Query Match 51.5%; Score 3841.5; DB 2; Length 1454;
Best Local Similarity 53.1%; Pred. No. 1.8e-153;
Matches 768; Conservative 208; Mismatches 411; Indels 59; Gaps 21;

QY 6 RYKHVKVKKHWTVAVASGLITLITL-----GSSVSAETEQQTSDDKVVTQK 53
DB 11 RKMKYSGK--MLVIAGSVIIIGVTSFIQAQADVSKQGVVVVTVAVNQNSDATTTDK 67

QY 54 S--EDDKA--ASESSQTDAPKQAQTEQTOAQSOA--NVADTSTSTKTPSPQNIITQAN 108
DB 68 SITTPKATTADTSTND---KATTADTSTNDKATTTADTSTN-DKATTTADISTNNK 122

QY 109 SDDKVTNNTKSEBQTSBRTKQSEBQTTASSQALTOAKAELTKQRTQAAQENKPNVDL 168
DB 123 ATTADTSTNNKATTAD--TSDNNSATTSKDVSSLA-----QKSAIDNNSKTTDT 174

QY 169 AA-----IPNVKIDGKYIIGSDQPKKFNALTNNKVLVFDKNTGALTDTTSQYQKQG 223
DB 175 AASFEASSKNLTKIDKTYIYDNGQIKNFATVDGKLVFDKDTGALADTNDIQFLEG 234

QY 224 LTKLNNDYTPHQINQVNFENSTLETINDYVADSWYRPDKILKNGKWTWASSSDRPLLM 283
DB 235 LFSNNNYTEHNASVGTSSASTYNTVDGYLTADSWYRPDKIFVNGQWNESSKDDDLRPLLM 294

QY 284 SWMPDKQTOIAYLNYMNQ--QGLGTGENTYADSSQESLNLAQTVQVKIETKISQTOQTOW 342
DB 295 TWMPDKATHVNLNWKYLDATETVYTSDDSDQDALNKAQNIQVKEIKISQEVQOTOW 354

QY 343 LRDIIINSFYKTOPNNSQTSDESAGEKDHLOGGALLYNSDKTAYANSDYRLNRPPTS 402
DB 355 LKDDISKFVDSQSNWNIASESGT---DHLGGALLYNSDKTPDANSDYRLNRPFTN 410

QY 403 QTGPKPYFDNSSGGYDFLLANDINDSNPNVQAEQLNWLHYLMYVGSIVANDPEANFGV 462
DB 411 QTGPLYTTDPTQGGYDFLLANDVNSNPVQAEQLNGMYLLNFGSITNNDADANFDSI 470

QY 463 RVDADVNNADLQIATSDYLKAHYGVDSKSKNAINHLSTLEAWSNDPOYNKDTKGAQLP 522
DB 471 RVDADVNDADLQIATADYFKAAYGVDSKSDAISNQHVSILEWSDNDADAEVYKNDGNQLS 530

QY 523 IDNKLRLSLLYALTRPLEKDAKNKEIRSGLEPVIITNSLNRSAAEGKSNRMANYIFIRA 582
DB 531 MDNKLRLSLKYSLTWPPVDHYGNK---RSGLEPFLTNSLVNRTNDSRDTAQNVSFVRA 587

QY 583 HDSEVQTVIAKIIKAQINPKTDLFTTLDLKOAFKINYNEDMRQAKKYQTSNIPATYAL 642
DB 588 HDSEVQTVIAEIIKQIRIDPSDGLSPTMDQLTEAFKINADQLTKDBFTQNIPTSYAT 647

QY 643 MLSNKDSITRLYYGDMYSDGQYMATKSPYYDAIDTLKARIYAAGQDMKITVYEGDK 702
DB 648 ILTNKDTVPVVYGYMDYTDGQYMATKSLYYDAIDTLKSRIKYVSGGQTMGMYKMGDS 707

QY 703 SHMDWDYTGVLTSVRYGTGANEAATDQSEATKQGMVITTSNNPSSLKLNQNDKIVNMG 762
DB 708 SWAADSRYGILTSVRYGNGAMTATAGTNETHGLAVIESNPNDLKUSQDQVVDWMI 767

QY 763 AHKNOEYRPLLLTTKDLGTSYTSDA--AAKSLYRKTNDKGLVDFDASDIQGYLNPQVSGYL 821

DB 768 AHKNOYRPAALLTKDIDITVSDSVQSOLIRYTNNSGQLIFNSSDIVGTANPQASGYL 827
QY 822 AVWVPVGASDNQDVRVAASNKANATGOVYESSALDSOLIEYGFSPQDFVTKDSVYTNK 881
DB 828 AVWVPVGASDQTDARTESSTATTDDQTLHSNAALDSQVYFESFSPQSTPTTEAEYANV 887
QY 882 KIAQNVQLFKSGMGTSPFEMAPQVYSSDEGSFLDSIIQNGYAFEDRYDLAMSKNKYSSQ 941
DB 888 QIANNTDLYKSWGITNFEPPQVRSSTSDSFLDSIIQNGYAFEDRYDLGNTPTPKYGTVD 947
QY 942 DMINAVKALHKSIGIQUIADWPDQIYNLPKQEVVTVATRVNDYGEYKDSIKNTLYAANT 1001
DB 948 QLRTAILKALHATGIKAMADWPDQIYNLTCKEYAVQVRVNSNGIYNQDSVINKTLVASQT 1007
QY 1002 KSNQKDYQAKYGGAPLSLAALKYPSIFNRTQISNGKKIDPSEKITTAKAKYFNGTNIIGR 1061
DB 1008 VGGG-EYQALYGEFDEIKLPSLPEKQISTGVPMASEKIKWSAKYFNGTNIQGR 1066
QY 1062 GVGYYLKNASDKYFEL---KGNQTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNS 1118
DB 1067 GAYVYLKQWATNEYFKVSTSSNSVFLPKQLTNEESNTGFISTDGWTYISTSGYQAKDT 1126
QY 1119 FVQDAKGNYYFDNNGHWYGLQNLGEVOYFSLNSGVQLRESFLENADGSKNYFGHLGNR 1178
DB 1127 FIQDDKSNYYFDPKNGYMTYGFQTVNDNNYYFLPNGIELQDAILEDKSGDVYFENQYKQ 1186
QY 1179 YSNGYYSFNDKSWRYFDASGYMA-VGLAKTI---NGNTQYFDDQGVYVKGAMITGSDGK 1233
DB 1187 TIDGYTMLANKT-WRYFDKNGVWANNAGLTVTVDGQKHQYDPKNGIQVKGTSVVKADGK 1245
QY 1234 KRYFDDGSGNMANVRPANDKNGDWYLLNSDGIALVGQVQTINGKTYIFGQDGKQIKGKIIT 1293
DB 1246 LAYFDTDSCEWTVNRFGEKTDGTSYFGADGLAVTGAQITISGKLPFDADGQIKGKEAT 1305
QY 1294 D-NGKUKYPLANSBELARNIPATDSNNWYFSGSDGAVTGSQTIAGKKLYFPASDGKQVK 1352
DB 1306 DKKGKMHYYDADSGEMTTNRFEKLSGDSWMYFNKNGVVTGAQVINGQHLFFESNGNQVK 1365
QY 1353 G-SFVYNGKHYHYHADSGELQVNRPEADKGNWYLLNSGEALTSQSRINDQRFVFTRE 1411
DB 1366 GREYTATDGKMYRYDADSGDMVTNRPFRISDSGWAYFGANGVAVTGAQNINGQOLYFDTK 1425
QY 1412 GKQVKG 1417
DB 1426 GHQVKG 1431

RESULT 6
Q55265 STRSL
ID Q55265_STRSL PRELIMINARY; PRT; 1577 AA.
AC Q55265
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Glucosyltransferase precursor.
GN Name=gtfm;
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=95122197; PubMed=7822030;
RA Simpson C.B., Giffard P.M., Jacques N.A.;
RT "Streptococcus salivarius ATCC 25375 possesses at least two genes
RL coding for primer-independent glucosyltransferases.";
RL Infect. Immun. 63:609-621 (1995).
DR EMBL; L35928; AAC41413.1; -; Genomic_DNA.
DR PIR; T30858; T30858.
DR HSSP; P06653; 1H8G.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.


```
DR InterPro; IPR004829; Csurface_antigen.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 4.
DR Pfam; PF02324; Glyco_hydro_70; 1.
DR ProDom; PD153432; Csurface_antigen; 1.
KW Signal; Transferase.
FT SIGNAL 1 38 Potential.
FT CHAIN 39 1577 Glucosyltransferase.
SQ SEQUENCE 1577 AA; 175290 MW; 3EFB898A7D3A7BF3 CRC64;

Query Match 50.8%; Score 3788.5; DB 2; Length 1577;
Best Local Similarity 49.1%; Pred. No. 3.4e-151;
Matches 777; Conservative 208; Mismatches 412; Indels 185; Gaps 21;

QY 1 METKRYKMHKVKCHWTVAVAS-----G 24
   |||:|||||:|||||:|
Db 1 MENKVRPKLHKVKKWVTIGVTLSNVALAGSLLAQKVEADETSAPNGDGLQQLSEDG 60
   |||:|||||:|||||:|
QY 25 LITLGT-----LGSVSAETEQTSDKVTQKSEDDKAASESSQTDAP 69
   |||:|||||:|||||:|
Db 61 TASLVTTTTVEQASQAQSVASVATASVSHETSFOAATSVAQEAQAQTSFVASQEVA 120
   |||:|||||:|||||:|
QY 70 KTKAQGT--EQTAQSQANVADTSITKETPSQ----- 101
   |||:|||||:|||||:|
Db 121 VSSQTQSSQGTQTTBEQVSGQTSTQVAGQTSAPVTEQARPRVLNAAPAIATRAA 180
   |||:|||||:|||||:|
QY 102 -----NITQAAASDDKTVT-----NTKSEEAQTSBERTKQSEEAQTASSQAL 144
   |||:|||||:|||||:|
Db 181 DSTIRINANRNTITITAGTTPNVTIIGPNTPKENVTVTPNGRPNVTIVTQPNQN 240
   |||:|||||:|||||:|
QY 145 TQAKAELTKQRTAAQAKNPNVDLAAI-PNVKQIDGKYIYIGSDGPKPKNFALTVNKKVL 203
   |||:|||||:|||||:|
Db 241 KPVPQSPQPNKVPQNPQSLDYKPVASNLKTIQCKYV-ENGVVKKNAATELDGRLY 299
   |||:|||||:|||||:|
QY 204 YFDKNTGALTDTSTQYQKQGLTKLNDYTPHQINVFNTSLETIDNYTADSWRPKDI 263
   |||:|||||:|||||:|
Db 300 YFDE-TGAMVDQSKPIYRADAIPNNSIYAVYQAYDTSSKSFELNFLTADSWRPKQI 358
   |||:|||||:|||||:|
QY 264 LKNGKWTWASSDRLPLMSWMPDKQTOIAYLNVMNQGLGTGENYTADSSQESLNLA 323
   |||:|||||:|||||:|
Db 359 LKDGKWTWASTKDYRPLMTWMPDKVTQVNYLYNMSQQFG-NKTYTTDMKSYDLAAA 417
   |||:|||||:|||||:|
QY 324 QTVQVKIETKISQTOQTLWRLDIINSFVKTPQPNWSQTESDTSAGKDKHLQGGALLYNS 383
   |||:|||||:|||||:|
Db 418 ETVQGIETIERIGREGNTTWLRQLMSDFIKTPQGNWSEEDNLLVG-KDHLQGGALTF 476
   |||:|||||:|||||:|
QY 384 DKTAYANSYRLNRTPTQTKPKPKYFEDNSSGGYDFLLANDIDNSNPVVAQBLNWLHY 443
   |||:|||||:|||||:|
Db 477 SATSHANSDFRLMNRPTTQGTGRKYHIDRSNGYELLANDIDNSNPVVAQBLNWLHY 536
   |||:|||||:|||||:|
QY 444 LMNYGIVANDPEANFDGVRVADVNDVNADLLQIASDYLKAHYGVDKSEKNAINHLSILE 503
   |||:|||||:|||||:|
Db 537 IMWIGSILGNDPSANFDGVRIDAVNDVDLQIASDYFKEKRYVADNEANAIAHLSILE 596
   |||:|||||:|||||:|
QY 504 AMSDNDPQYNKDTKGALPDKRLSLLYALTRPLEKDASNKNEIRSGLEPVITSLNN 563
   |||:|||||:|||||:|
Db 597 AWSYNDHQYNKDTKGALSDINPLRETLTTPLR-----KSNRYGSLERVITSLNN 648
   |||:|||||:|||||:|
QY 564 RSABGNKSERMANYIIRAHDSVQVVIKIIKAQINPKTDGLTFTLDELKQAFKYNED 623
   |||:|||||:|||||:|
Db 649 RSSEQHTPRDANYIIFVRAHDSVQAVLANIISKQINPKTDGFTFTMDLQKAFEYINAD 708
   |||:|||||:|||||:|
QY 624 MRQAKKYYTQSNPTPAVALMSNKDSITRLYGDWYSDQVYATKSPYYDAIDTLLKAR 683
   |||:|||||:|||||:|
Db 709 IAKADKKYTOYNIIPAAVATWLTNKSITRYYGDLFTDDGOYMAEKSPYYNAIDALLRAR 768
   |||:|||||:|||||:|
QY 684 IKYAAGQDMKITYVEGDKSHMDWDYTGVLTSVRYGTGANEATDQGEATKTQGMVITS 743
   |||:|||||:|||||:|
Db 769 IKYVAGQDMKVTKLNGYE-----IMSSVRYKGAEANQLGTAETRNQGLVLTA 819
   |||:|||||:|||||:|
QY 744 NNPSLKLNDQKIVVNMGAHKQYRPLLLITTKDGLTSYTSDA-AAKSLYRKTNDKGL 802
   |||:|||||:|||||:|
```

```
Db 820 NRPDKMLGANDRLVNMGAHKQYRPLLLSKSTGLATLYLKDSDVPAGLVRYTDNQNL 879
QY 803 VFDASDIQGYLNPOVSGYLAVWVPVGASNDQVRVAASKANATQGVYESSALDSQLLY 862
   |||:|||||:|||||:|
Db 880 TFTADDIAGHSTVEVSGYLAVWVPVGASENQDARTKASSTKKE-QVFESSAALDSQVY 938
   |||:|||||:|||||:|
QY 863 EGFSNFQDPTVKDSDYTNKKTIAQNVQLFKSMGVTSPFEMAPQVYSSSEDGFLDSIIQNGYA 922
   |||:|||||:|||||:|
Db 939 EGFSNFQDPTVKPSQYTNRVIAQNAKLKFEWGITSFEPAPQVSSODGTFLODSIIENGVA 998
   |||:|||||:|||||:|
QY 923 FEDRVDLAMSNNKNTYGSQODMINAVKALHKSIGQIVADWVPDQIYNLPCKEVVTAIRVD 982
   |||:|||||:|||||:|
Db 999 FEDRVDIAMSNNKNTYGSGLKDLMDALRALHAEGISALADWVPDQIYNLPCKEVVTAIRVD 1058
   |||:|||||:|||||:|
QY 983 YGEYRKDSEIKWTLVAANTKSNKDYQAKYGGAPLSELAAKYPSIPNRTQISNGKIDPS 1042
   |||:|||||:|||||:|
Db 1059 YGTPRPAEINSLYAAKTRTFGNDFOGKYGGAFDELKAKYPAIFERVQISNGRKLTTN 1118
   |||:|||||:|||||:|
QY 1043 EKITAWKAKYFNGTILGRGVGYVLKDNASDKYFELKGNQTYLPKQMTNKEASTGFVNDG 1102
   |||:|||||:|||||:|
Db 1119 EKITQMSAKYFNGSIQGTGARYVIQDNATNQYFSVKAGQTFELPKQMTITGS-GPRRVG 1177
   |||:|||||:|||||:|
QY 1103 NGMTFYSTSGYQAKNSFVQDAKGNWYFDNNGHMVYGLQQLNGEVOYFSLNSGVQLRESFL 1162
   |||:|||||:|||||:|
Db 1178 DDVQYLSIGGYLAKNTFIQVGANQWYFDPKGNMWTGEQVIDGKKYFFLDNGLQLRHVLR 1237
   |||:|||||:|||||:|
QY 1163 ENADSGKNYFHLGNRYNSGYSPNDSSK-WRYFDASGVMAVGLKTINGNTQYFDO-DGY 1220
   |||:|||||:|||||:|
Db 1238 QGSDGHVYVYDPKGYQAFNGFYDFAGPRQDVRVYFDGNGQMYRGLHDMYGTTFYFDEKGTI 1297
   |||:|||||:|||||:|
QY 1221 QVKGAWITGSDCKKBYFDDGSGNMAVNRFA-NDKNGDWYLYNSDGTALVGVTINKTKYY 1279
   |||:|||||:|||||:|
Db 1298 QAKDFIIRFADGRTYFIPDTGNLAVNRPAQNPENKAWYLYDSNGYAVTGLQTINGKQYY 1357
   |||:|||||:|||||:|
QY 1280 FQDQGGKQIKGIITONGKLYEL-ANSGLARNIFATDSQ----- 1318
   |||:|||||:|||||:|
Db 1358 FVNEGRQVKGHFVTINQ-RYFLDGDGSEIAPSRFTENNKNWYVDGNGKLVKGAQVING 1416
   |||:|||||:|||||:|
QY 1319 -----NNWYFSGDGVAVTGSQTIAG 1339
   |||:|||||:|||||:|
Db 1417 NHYFENNDSYQVKGAWANGRYYDGSQAVSNQFIQIAANQAWYLNQDGHKVTGLQVINN 1476
   |||:|||||:|||||:|
QY 1340 KKLVPASDCKOVKGSFVTYNGKVHYHADSGELQVNRFEADKGNWYLYDSNGEALTGSQ 1399
   |||:|||||:|||||:|
Db 1477 KYTYFGSNGAQKGLLTVQGGKKYFPAHTGSGQVNRVFEAARGCWYFNSAGQAVTGOQ 1536
   |||:|||||:|||||:|
QY 1400 RINDORVFFTEGKQVKGDVAY 1421
   |||:|||||:|||||:|
Db 1537 VINGQLYFDGSGRQVKGRVY 1558
   |||:|||||:|||||:|
```

RESULT 7

```
GTFC_STRMU STANDARD; PRT: 1455 AA.
ID GTFC_STRMU STANDARD; PRT: 1455 AA.
AC P13470; O69382; O69385; O69388; O69391; O69397; P05427;
DT 01-NOV-1988 (Rel. 09, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Glucosyltransferase-SI precursor (BC 2.4.1.5) (GTF-SI)
DE (Dextranucrase) (Sucrose 6-glucosyltransferase).
GN Name=gTfC; OrderedlocusNames=SMU.1005;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=GS-5;
RX MEDLINE=89137980; PubMed=2976010; DOI=10.1016/0378-1119(88)90382-4;
RA Ueda S., Shirota T., Kuramitsu H.K.;
RT "Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.";
RL Gene 69:101-109(1988).
RN [2]
```

RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MT4239 / Serotype c, MT4245 / Serotype e, MT4251 / Serotype f,
RC MT4467 / Serotype e, and MT8148 / Serotype c;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans";
RL FEMS Microbiol Lett. 161:331-336(1998).
RN [3]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=UA159 / ATCC 700610 / Serotype c;
RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;
RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,
RA Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
RT pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
RN [4]
RP NUCLEOTIDE SEQUENCE OF 1-349.
RC STRAIN=GS-5;
RX MEDLINE=87308013; PubMed=3040685;
RA Shiroza T., Ueda S., Kuramitsu H.K.;
RT "Sequence analysis of the gtfB gene from Streptococcus mutans.";
RL J. Bacteriol. 169:4263-4270(1987).
CC -!- FUNCTION: Production of extracellular glucans, that are thought to
CC play a key role in the development of the dental plaque because of
CC their ability to adhere to smooth surfaces and mediate the
CC aggregation of bacterial cells and food debris.
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-
CC fructose + (1,6-alpha-D-glucosyl) (n+1).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
CC 1,3-linked glucose and some 1,6 linkages). GTF-SI synthesizes both
CC water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
CC forms of glucans.
CC -!- SIMILARITY: Belongs to the glucosyl hydrolase 70 family.
CC -!- SIMILARITY: Contains 5 cell wall binding repeats.
CC
CC This Swiss-Prot entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL; M22054; AAA88592.1; -; Genomic DNA.
CC EMBL; D88652; BAA26102.1; -; Genomic DNA.
CC EMBL; D88655; BAA26106.1; -; Genomic DNA.
CC EMBL; D88658; BAA26110.1; -; Genomic DNA.
CC EMBL; D88661; BAA26114.1; -; Genomic DNA.
CC EMBL; D89978; BAA26120.1; -; Genomic DNA.
CC EMBL; AE014940; AAN58706.1; -; Genomic DNA.
CC EMBL; M17361; AAA88589.1; -; Genomic DNA.
CC PIR; JTO345; JTO345.
CC HSSP; P06653; 1H8C.
CC InterPro; IPR002479; Cell_wall_bd_put.
CC InterPro; IPR003318; Glyco_hydro_70.
CC Pfam; PF01473; CW_binding_1; 2.
CC Pfam; PF02324; Glyco_hydro_70; 1.
CC Complete proteome; Dental Caries; Glycosyltransferase; Repeat; Signal;
CC Transferrase.
CC
CC SIGNAL 1 34
CC CHAIN 35 1455 Glucosyltransferase-SI.
CC REPEAT 1126 1159 A repeat.
CC REPEAT 1169 1200 A repeat.
CC REPEAT 1227 1238 C repeat.
CC REPEAT 1253 1303 AC repeat.
CC REPEAT 1318 1330 A repeat (incomplete).
CC REGION 35 1050 Catalytic (approximate).
CC REGION 1126 1455 2.4 A, 1 C and 1 AC repeats.
CC REGION 1126 1455 Glucan-binding (approximate).

FT VARIANT 21 21 V -> I (in strain GS-5).
FT VARIANT 81 81 P -> L (in strain MT4239).
FT VARIANT 106 106 D -> V (in strain GS-5).
FT VARIANT 116 116 S -> A (in strain GS-5 and strain
MT4467).
FT VARIANT 126 126 A -> T (in strain GS-5).
FT VARIANT 150 151 SR -> PK (in strain GS-5, strain MT4239
and strain MT4467).
FT VARIANT 256 256 A -> V (in strain GS-5 and strain
MT4467).
FT VARIANT 425 425 R -> N (in strain MT4251).
FT VARIANT 519 519 Y -> D (in strain MT4245 and strain
MT4251).
FT VARIANT 538 538 R -> K (in strain MT4245 and strain
MT4251).
FT VARIANT 545 545 Y -> F (in strain MT4245 and strain
MT4251).
FT VARIANT 597 597 N -> D (in strain MT4245, strain MT4251,
strain MT4467 and strain MT8148).
FT VARIANT 600 600 R -> K (in strain MT4245, strain MT4251,
strain MT4467 and strain MT8148).
FT VARIANT 601 601 A -> T (in strain GS-5).
FT VARIANT 614 614 M -> T (in strain GS-5).
FT VARIANT 727 727 T -> I (in strain MT8148).
FT VARIANT 734 734 A -> V (in strain MT8148).
FT VARIANT 964 964 L -> F (in strain MT4239).
FT VARIANT 1113 1113 N -> Y (in strain MT4239).
FT VARIANT 1118 1118 A -> T (in strain MT4239).
FT VARIANT 1204 1204 I -> V (in strain GS-5, strain MT4239,
strain MT4467 and strain MT8148).
FT VARIANT 1208 1208 V -> I (in strain MT8148).
FT VARIANT 1292 1294 DGH -> NGV (in strain GS-5, strain MT4467
and strain MT8148).
FT VARIANT 1305 1369 Missing (in strain MT4245).
FT VARIANT 1326 1326 I -> V (in strain MT8148).
FT VARIANT 1331 1331 T -> A (in strain GS-5, strain MT4239,
strain MT4467 and strain MT8148).
FT VARIANT 1377 1377 R -> K (in strain MT8148).
FT VARIANT 1398 1398 V -> I (in strain MT8148).
FT VARIANT 1424 1424 D -> N (in strain MT4239).
FT VARIANT 1439 1439 V -> I (in strain MT4239 and strain
MT8148).
FT VARIANT 1444 1444 S -> P (in strain MT8148).
FT CONFLICT 1337 1455 ORLYKNGVQAKGELITERRKRIKYDDNSGNEVRNVYR
TSSGNWYFGNDGYALIGHVVEGRVYDENGVYASHD
QRNWDYFRRDFGRGSSAVFRHSRNGFFDNFFRP ->
HASILSLMVFRLRESLSQSVKVVNTMILIPKMFVIVM
(in Ref. 1).
SQ SEQUENCE 1455 AA; 162966 MW; 3CB455A994FEC86 CRC64;
Query Match 50.2%; Score 3743; DB 1; Length 1455;
Best Local Similarity 51.1%; Pred. No. 2.5e-149;
Matches 753; Conservative 232; Mismatches 384; Indels 106; Gaps 24;
QY 1 METKRYKMKVKKHWTVAVASGLITLTGSSVSASAEQTQSDKVVTKSEDDKAA 60
DB 1 MEKKYRFLKRVKRWTVSVASAVTL-TSSGSLVRADS-----TDRQQA 47
QY 61 SESSQTDAPKTQAQTEQTAQSOANVADTSTSTTKETPSQNTTQANSDDKTVNTKSE 120
DB 48 VTESQASLVTTSEA-AKETLTATDTSTATSQPTATVTDVNVSTTNQS-----TNTAN 101
QY 121 EAQTSEERTKQSEEAQTASSQALQAQA--ELT----- 152
DB 102 TANFDVKPTTTTSEQSKTDNSDKIIATSKAVNRLTATGKFPVANNNTAHSRTVTDKIVPIK 161
QY 153 -----KQRTAAQENKPNVDLAAIPNVK---OIDGKYYYIGSDGPKQKQNFALTQNVKVL 203
DB 162 PKIGKLPSSLSQD-----DPAALGNVKNIRKVKNGKYKEDGLQKNYALNNGKTF 216
QY 204 YPDKNKTGALTDTSTQYQFGKGLTKLND-----YTPNQIVNFENTISLETIDNVYVADSWSYR 259
DB 217 FEDE-TGALSNNLPSKGNIT--NNDNTNSPAQYNQVYSTDAANFEHVDHVLTAESWYR 273

Db 413 QGALLYNNDSRTSHANSYRLNLRPTTSQTGKRNPKYTKDTSNGGFEFLANDINSNP 472
QY 432 VVOAQLNLWLHYLMNYGSIIVANDPEANFDGVRVDVAVNNVADLQIASDYLKAHYGVDKS 491
Db 473 AVQAEQLNLWLHYIMNIGTITGSGSEDFDGVDAVDVAVNNVADLQIASDYFKAKYGADQS 532
QY 492 EKVAINHLSLEAWSNDPOYNKDTKGAQPIPNKURLSLLYALTALPLEKANSKNEIRS 551
Db 533 QDAIKHLSLEAWSHNDAYNEDTKGAQLPMDPMELALVYSLRPIGN-----RS 584
QY 552 GLEPVTNSLNNRASAEGKSERMANYIFIRAHSEVQTVIAKIIKQINPKTDLGTLFTLD 611
Db 585 GVEPLNSLNDRESGSKRMANYAFVRAHSEVQSIIGQIIKNEINPSTGNTFTLD 644
QY 612 ELKQAFKYNEDMRQAKKYTQSNIPATYALMLSNKDSITRLYGYDMYSDDGYMATKSP 671
Db 645 EMKAFEIYNKDMRSANKQYQYNIPSAVALMLTKDVTFRVYGYDMYTDGQYMAQKSP 704
QY 672 YYDAIDTLKARIKYAAGGDMKITVYEGDKSHMDWDYTGVLTSVRYGTGANEAATOGSE 731
Db 705 YYDAIETLLKGRIRYAAGGDMKVNI-GYGTNGWDAGVLTSVRYGTGANSASDTGTA 763
QY 732 ATKTQGMVITTSNNPSLKLNDQKVIIVNMGAAHKNQYRELLLTTKDGLTSYTSDAAKS 791
Db 764 ETRNQGMVIVSNQPALRTSN--LTINMGAAHRNQAYRELLLTNDGVATYLNDSANG 821
QY 792 LYRKTNDKGBELVPDASDIQGLNLPQVSGVLAVPVPGASNDQVRVAASNKANATQVYE 851
Db 822 IVKYTDGNGNLTFSEANEIRGNPQVDGYLAVVPVPGASENQDVRVAPSKENSSGLVYE 881
QY 852 SSSALDSQLIYEGFSNFPQVTKDSYTNKIIAQNVLFPKSWGVTSFEMAPQVSVESDGS 911
Db 882 SNAALDSQVIYEGFSNFPQVFPNPSQYTNKIIAENANLFKSMGITSFEPAPQVSVSDGS 941
QY 912 FLDSIIQNGVAPEDRYDLAMSKNKYGSQODMINAVKALHKSIOVIADVDPDIYNLPG 971
Db 942 FLDSVLQNGVAFTRDYDGHKNDKYGSLADLKAALKSLHAVGISAIADVDPDIYNLPG 1001
QY 972 KEVVTATRVNDYGEYRKDSIKNTLYAANTKSNKGQYQAKYGGAFLSLAAYPSIFNRT 1031
Db 1002 DEVVTATRVNNGYGETKDAIDHSLYAARTFPGNDYQGYGAFIDELKRLYQIFDVR 1061
QY 1032 QISNGKKIDPSEKITAWKAFNGTNILGRGVYVLKDNASDKYFELKGNQTYLPKQMTN 1091
Db 1062 QISTGKRMTTDEKITKWSAKYMGNTNILDGSEYVLK-NGLNGYYGTNGKVSPLPKVGS 1120
QY 1092 KEASTGFVNDGNG-----MTFYSTSGYQAKNSFVQDAGKNWYFYDNNGHMYGLQ 1141
Db 1121 NOSTNGDNQNGDGSKGFKELFSVRYRNNQYAKNAFTKDNNGVYFYDNGSRMAVGK 1180
QY 1142 QLNGEVQYFSLNGVQLRESFLENADGSKNYFGLHGNRYSGYV--SFNDNSK----- 1191
Db 1181 TIDGKYFFLANGVQLRDGYRQNRQOVYFYDQNGVLSANGQDPPKPDNNNTSGNRQV 1240
QY 1192 -----WRYFDASGVMAVLKTINGNTQYFDQGYQVKGAWITGSDGKKRYFDDGSGNMA 1245
Db 1241 QIGNNWAYYDNGKRVGHONINGQELFFDNNGVQVKGRTV-NENGALRYIDANSGEWA 1299
QY 1246 VNRFPANDKNGDWYLLNSDGIALVGVTINGKTYFYFGQDQKQIKGIITDNGKLYPLANS 1305
Db 1300 RNRFAIEFGWAYFNNNDTAVKGSQNIINGQDLYFDQNGRQVKGALANVDGNLRYDYVNS 1359
QY 1306 GELARNIPATDSQNNWYFSGSDGVAVTGQTAGKLYFASDGKQVKGFSVYNGKVHY 1365
Db 1360 GELYNRFP-HEIDGSGWYFDGNGNAVKGMWNINGQMLLFDNNKGQIKGHLVRVNGVRYF 1418
QY 1366 HADSGELQVNRFEADKGNWYLDNSG 1392
Db 1419 DPNSGEMAVNRVVEVSPGWWYFDGEG 1445

RESULT 9
Q55264_STRSL

ID Q55264_STRSL PRELIMINARY; PRT; 1449 AA.
AC Q55264;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Glucosyltransferase precursor.
DE Name=gtfII;
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
OX [1]
RN NUCLEOTIDE SEQUENCE.
RX MEDLINE=95122197; PubMed=7822030;
RA Simpson C.L., Giffard P.M., Jacques N.A.;
RT "Streptococcus salivarius ATCC 25975 possesses at least two genes
coding for primer-independent glucosyltransferases.";
RL Infect. Immun. 63:609-621(1995).
DR EMBL; LJ5495; AAC1412.1; -; Genomic_DNA.
DR FIR; T30857; T30857.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Signal; Transferase.
FT SIGNAL 1 35 Potential
FT CHAIN 36 1449 Glucosyltransferase.
SQ SEQUENCE 1449 AA; 159984 MW; DD62F07306E86A46 CRC64;
Query Match 49.8%; Score 3713.5; DB 2; Length 1449;
Best Local Similarity 51.4%; Pred. No. 4.4e-148;
Matches 754; Conservative 207; Mismatches 409; Indels 97; Gaps 22;

QY 1 METKRYKMHKVKHVVAV-----ASGLITLT-----TTLGSSVSAET 41
Db 1 MDKRVHYKMHKVKQWTTIAVTGLSGLAVSAVSLGTDGVDGVADEHTDATVAIPDITVD 60
QY 42 EQGTSKVVYTKSEDDKAESSQTD--APKTKQATQEQTAQSQANVADTSTSI----- 94
Db 61 GTVSNDDTTAAQDPTTAAVATNDVATDQATPTATFDLTDTTNTVAANAVDTVATGTDRA 120
QY 95 --TKETPSQNIIT--QANSDDKTVNTKSEEAQTSBEERTKQSEEAQTASSQALTQAKAE 150
Db 121 ATNTDTATNDTAVDTTNNNTTDTTVDRAATERRATGARGPT--GGRATPVGN 178
QY 151 LTKQRQTAQENKNPVDLAAIPNV-----KQIDGKYIYIGSDGQPKKFNALTNNKV 202
Db 179 TNNANNTVTVNN---DLPATNNVTDGSPSHIKTINGKQYVVEDDGTIRKNVYLERIGGS 235
QY 203 LYFDKNTGALTSTQYQF-KQGLTKLNNDYTPHNQIVN-----FENTS---LETINDYVT 253
Db 236 QYFNAETGELSNOKEYRFDKNGTGSSADSTNTNTVNGDKNAFYGTDDKIDELVDGYFT 295
QY 254 ADSWYRPDKILKNGKTTWASSSDLRPLLMWSWPKQTOIAYLYNMNQGLGTGENYATD 313
Db 296 ANTWTRPEILKGEWASTASTDKRPLLTVMWPSKAIQASLYNMYKEQGLGTNYQYTSF 355
QY 314 SSQESLNLAAQTVQVKIETKISQTOQTOWLRDIINSFVKTPQNNWSQTSDETSAGSKHL 373
Db 356 SSQTQMDQAALEBVQKRIERIAREGNTDWRITTKNFVKTPQGNWSTSE--NLDNDHL 412
QY 374 QGALLYNSDKTAYANSDYRLNLTPTTSQTK--PKTFEDNSSGGYDFLLANDINSNP 431
Db 413 QGALLYNNDSRTSHANSYRLNLRPTTSQTGKRNPKYTKDTSNGGFEFLANDINSNP 472
QY 432 VVOAQLNLWLHYLMNYGSIIVANDPEANFDGVRVDVAVNNVADLQIASDYLKAHYGVDKS 491
Db 473 AVQAEQLNLWLHYIMNIGTITGSGSEDFDGVDAVDVAVNNVADLQIASDYFKAKYGADQS 532
QY 492 EKVAINHLSLEAWSNDPOYNKDTKGAQPLIPNKLRLSLLYALTALPLEKANSKNEIRS 551

Db 533 QDOAIKHLSTLEAWSHNDAYNEDTKGAQLPMDPMHLALVYSLLRPIGN-----RS 584
Qy 552 GLEPVTITSLNNRSABGKNSERMANYIFIRAHSEVOTVIAKIIKAQINPKTDGLTFTLTD 611
Db 585 GVEPLISNLSNDRSESGKSKRMANYAFVRAHSEVQSIIGQIIKNEINPQSTGNTFTLD 644
Qy 612 ELKQAFKINEDMROAKKYTQSNIPITAVALMSNKDSITRLYYGDMYSDDGQYWKSP 671
Db 645 EMKKAPEIYNKMRSAKQYTOYNIPSAVALMLTHKDTVPVRYGDMYTDGQYMAQKSP 704
Qy 672 YYDAITLKLARIKYAAGGDMKITVYEGDKSHMDWDTGVLTSVRYGTGANATDQGSSE 731
Db 705 YYDAITLKLGRIRYAAGGDMKVNYI-GYGTNGHWDAGVLTSVRYGTGANSASDTGTA 763
Qy 732 ATKQGMVITSNNSPLKLNQNDKVIIVNMAAHKQNEYRPLLTTKDGLTSYTSDDAAKS 791
Db 764 ETRNQGMVITSNNSPLKLNQNDKVIIVNMAAHKQNEYRPLLTTKDGLTSYTSDDAAKS 821
Qy 792 LYRKTNDKGLVDPASDIQGYLNPQVSGYLAVWVPVPGASDNQDVRVAASNKANATQVYE 851
Db 822 IVKYTDGNGNLTFSAEIRGIRNPQVDGYLAVWVPVPGASENQDVRVAPSKERNSSGLVYE 881
Qy 852 SSSALDSQLIYEGFSNFQDFVTKDSYTNKKIAQNVQLFKSMGVTISFEMAPQVSSDDGS 911
Db 882 SNAALDSQVIYEGFSNFQDFVQNPQSYTNKKIAENANLFSWGITSFEFAPQVSSDDGS 941
Qy 912 FLDSIIQNGYAFEDRYDLAWSKNNKYGSQODMINAVKALHKSQIOVIADWVPDQIYNLPG 971
Db 942 FLDSVIQNGYAFEDRYDLAWSKNNKYGSQODMINAVKALHKSQIOVIADWVPDQIYNLPG 1001
Qy 972 KEVVTATRVNDYGEYKDSIEKNTLYAANTKSNKDYQAKYGGAFSELAAPYSPFNKT 1031
Db 1002 DEVVTATRVNNGYETKDGAIIDHSLYAAKTRTFGNDYQGYGAFGLDELKRLYPOIFDRV 1061
Qy 1032 QISNGKKIDPSEKITAUKAKYFNGTNIILGRGVYVLKDNASDKYFELKGNQTVLPKQMTN 1091
Db 1062 QISTGKRWTTDEKLTQWAKYMGNTNIDRGSEYVLK-NGLNGYYGTNGKGLPKVVGGS 1120
Qy 1092 KEASTGFVNDGNG-----MTFYSTSGYQAKNSFVQDAGKNNYFYDNNGHMYYGLQ 1141
Db 1121 NQSTNGDNQNGDGGKFEKELFSVRYRYNNGYAKNAFIKNDNGVNYFYDNNSGRMVAGEK 1180
Qy 1142 QLNGEVQYFLSNGVOLRESFLENADGSKYFHLGNRYNGYY--SFDNDSK----- 1191
Db 1181 TIDGKYFFFLANGVQLRDGYRQNRGQVFFYDQNGVLNANGKQDPKPDNNNNSAGRNQFV 1240
Qy 1192 -----WRYFDASGVMAVLKTINGNTQVFDQGYQVKGAWITGSDGKKRYFDDGSGNMA 1245
Db 1241 QIGNWYWAYDNGKRVGTGHONINGQELFFDNGVQVKGRTV-NENGALRYDYDANGGWA 1299
Qy 1246 VNRFPANDKNGDWYLLNSDGIALVGVQTINGKTYFFQDQKQIKGKIITDNGKLYFLANS 1305
Db 1300 RNRFAETEPGWAYFNNDGTAFAVGQSONINGQDLYFDQNGRQVKGALANVDGNLRYDYVNS 1359
Qy 1306 GELARINIFATDSQNNWYFSGDGVAVTGSQTIAGKLYFASDCGVKSGSVTVNGKVHY 1365
Db 1360 GELYRNRFP-HEIDGWSYTFDNGNAVKGWYNGVINGQNLFFDNGKQIKGHLVRVNGVRYF 1418
Qy 1366 HADSGELQVNRFEADKDGNNYYLDSNG 1392
Db 1419 DPNSGEMANVRWEVSPGWVYFDGEG 1445

RESULT 10

Q9WXJ5_9STRE
ID Q9WXJ5_9STRE PRELIMINARY; PRT; 1512 AA.
AC Q9WXJ5_9STRE
DT 01-NOV-1999 (T-EMBLrel. 12, Created)
DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE GTF-S.
GN Names=gtfst;
OS Streptococcus criceti.

OG Plasmid pAM1.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OX Streptococcus.
RN NCBI_TaxID=1333;
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=HS-6;
RA Inoue M., Fukui K., Miyagi A.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB026123; BAA77237.1; -; Genomic_DNA.
DR HSSP; P06654; 1MPE.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Plasmid.
SQ SEQUENCE 1512 AA; 167145 MW; 4C03D9CBC601FC14 CRC64;

Query Match 49.1%; Score 3661; DB 2; Length 1512;
Best Local Similarity 49.3%; Pred. No. 7.6e-146;
Matches 759; Conservative 230; Mismatches 400; Indels 150; Gaps 31;

Qy 1 METKRYKMHKVKCHWTVAVAS-GLI-----TLGTTTLGSSVAETEQTS--- 46
Db 1 MERKRYKLUHKVKQWTVIIVASAGLASVIGAGASQTVSADDINGASASAEQNTSASQ 60
Qy 47 DKVVTQ----KSEDDKAASESSQTDAPKTKQAQTEQTAQSQANVADTSTSIKTPSQN 102
Db 61 NKEVDSDAQAQATKATSEQAASVSDT-NGSETDQV-----NVDGVSANNQPEAPA-- 112
Qy 103 ITTOANSDDKTVNTKSEBAQTSERTKQSEBAQTTASSQALTOAKAELTKQRTAAQEN 162
Db 113 -APQAAAASNTNTAT-SEANTN---TAVSEAA--PAEENRTAEKAGADLSQDEEAAA--- 162
Qy 163 KNPVDLAAIPNVKQIDGKYIYIGSDGPKKNFALTNNKLVYFDKNTGALTDTSTQYQFQK 222
Db 163 -----LSLDNKKVGDGKYIYVADGSKYKNAFTVKGQLLYFDATGALSTSTYSFSQ 216
Qy 223 GLTKLNNDYTPHNQIVNFENTSLETIDYNTADSWYRPDKIILKNGKTWTASSEDRLRPL 282
Db 217 GLTPLVSDPSINKKAFDSSAKSFELDGYLTAAESWYRPTKILENGKTWVDSKETDLRPVL 276
Qy 283 MSWVPDKQTOIAYILNMQOGLTGENTYADSSQSLNLAQTVQVKIETKISQTOQTOW 342
Db 277 TSWVPDKQTOVAYLANYMS-KALGGKEBEFTTKTSQALNTAAEMIQMKIEQRISKEQTAW 335
Qy 343 LRDIINSFVKTOPNNNSQTESDTSAGEKDHLOGGALLYNSDKTAYANSDYRLLNRTPTS 402
Db 336 LRDAWAAFAVATOSRWNSDSE---QPKNDHLOGGALLYTNKLTWADSKYRLNRTPT 392
Qy 403 QTKPKPYFDNSGGYDFLLANDINDSNPVAQEQNLNHLHYLMNYGSIIVANDPEANPDGV 462
Db 393 QDGKTHYKADVEGVEFLANDVNSNPVQAEMLNQIHYLMWGSIVMGDKDANFDGI 452
Qy 463 RYDAVDNVDADLQTLASDYLKAHYGVDKSEKAINHLSILEAWSNDNDPOYNKDTKCAQLP 522
Db 453 RYDAVDNVDADTLQLTNTYFNAYGVDKSEQAQALAHISILEAWSYNDYNTQNTGGAALA 512
Qy 523 INKRLSLLYALTREPLEKDNASKNEIRSGLEPVTN--SLNRSAGKNSERMANYIFI 580
Db 513 MDNGLRLSLLYLTREPL-----SERTFGLSTLIKSEYGLTDRTKDKKYGDTQPSYFV 565
Qy 581 RAHDSVQTVIAKIIKAQINPKTDGLTFTLDELKQAFKIYNEDMRQAKKYYTQSTIPTAY 640
Db 566 RAHDSVQTVIAKIIKEKIDPTDGTFTLDQLKQAFDIYNKDMNSVEKHYTHYNIPAA 625
Qy 641 ALMLSNKDSITRLYYGDMYSDDGQYMAKSPYDAIDTLLKARIKYAAGQDMK----- 694
Db 626 AVMLSNMESVTRVYVYGDFTDQGYMETKSPYDAINTLLRARIYAAAGQTMKHKAYTP 685
Qy 695 --ITYVEGDKSHMDYDTGVLTSVRYGTGANATDQ--GSEATKTQGMVITSNPSIKLN 751

Db 686 SAAMKAKNPDSGSGVGNSEVLVSRFGQDVMSADMTGGQAKTSGMFSLIANNPELELD 745
Qy 752 QNDKVIIVMGAHKNQYRPLLLTTKGLTSTYTDAAAKSLYRKTNDKGLVDFASDIQ 811
Db 746 ANBEIKVNVGKIHAQAVRPLLLTTKGLQKYLNDSDT-NLTKVADKDGFTTFKGSIKG 804
Qy 812 YLNPQVSGYLAVWPVPGASNDQVRVAASNKANATG-QVYESSALDSQLIYEGFSNFQD 870
Db 805 YKQVEVNGYLSVWVPVPGAKSQDQIRVAASTKANAKGDKSYTASQALDSQLIYEGFSNFQD 864
Qy 871 FVTKDSYTNKIKANQVLPKSGWGTSEFMAPOVVSSEDSFLDSIIQNGYAFEDRDYLA 930
Db 865 FVQKDAQYTNKIKIAENTDLFKWGVTSFEMAPQVVSATDGFLLDSIIQNGYAFEDRDYLA 924
Qy 931 MSKNKYGSQODMINAVKALHSGIQVIADWVPDQIYNLPGKEVVVTRVNDYGEYRKDS 990
Db 925 MSKNKYGSKEDLANALKALHAAGIQALADWVPDQIYQLPGKEVVVTRVNDYGRVKIDQ 984
Qy 991 EIKNTLYAANTKSGKDYQAKYGAFLSELAAKYPSIPNRTQISNGKKIDPSEKITAWKA 1050
Db 985 PMVNLKYLANTKSSGKDFQAKYGGFEFLAELQKQYPEMFTAKMISTGKPIDSSVKLEWSA 1044
Qy 1051 KYFNGTNILGRGVYVLKDNASDKYFELKGNQTYLPLKQMT-NKEASTGTFVNDGNGMTFYS 1109
Db 1045 QYFNGTNVLRGTDTYLSDEGTGKYFTVNEKGEFLPAVLTDGKEAKTGFYNDGKGMTYFT 1104
Qy 1110 TSGYQAKNSFVQAKGNWYFDNNGHMVYGLQQLNGEVOYFSLNSGVOLRESFLENADGSK 1169
Db 1105 TAGSQAKSDFVTVAGNTYFDYTGHMVGTGNGINTKFFYFLPNGVWLKDAVWEDDRGRS 1163
Qy 1170 NYFGLHGRY-----SNGYISFDNDSK-----WRYFDASGVMAVGLKTINGNTQYFDQDGYQ 1221
Db 1164 VYTGKTVMYKGSNNNEFAM-TDSKGQLRFRHFDNYGFMVGLVTHGNVQYVDESGFQ 1222
Qy 1222 VKGANITSGDKKRYFPDGSNNMVRPANDKNGDWYILNSDGLALGVQVOTINGKTYFP- 1280
Db 1223 VKGDFVTDKAGQTRYFDKNTGNLVKQGFEN-QNGHWYISDDQGLIAKGAQTIKQKLYFD 1281
Qy 1281 GQDGKQIKGKIITD-----NG 1296
Db 1282 AKTGAQVKGDFVTDKQNTFFYSGDTCGLAVSTFFSTGNNAWFYADENGHVAKEKTING 1341
Qy 1297 KLKYFLANGSELARNIPATDSQ-----NNWYFSGDGVAVT 1332
Db 1342 QKLYFDTKTQQAQKFRVDAKGLRFDYADTALVTNSFLETRAGSNQWYMGADGVAVR 1401
Qy 1333 GSOTIAGKLYFP-ASDGKQVKGSEFT-YNGKVHYHYHDSGELQVNRFEADKQNGWYILDS 1390
Db 1402 GHQTIQSRHMYFPAETGQQAQKGIWVTDANGKRYFYDANTGDRVNVQFVL-VNGSWYFFGY 1460
Qy 1391 NGEALGTSORINQORVFFFTREGKQVKGDVAYDERRLLVY 1429
Db 1461 DGAAVTGFRDIRQCHLYFNPDPGTOAKGTTVKIDNRIYTF 1499

RESULT 11

Q56CX8_9STRE PRELIMINARY; PRT: 1506 AA.
AC Q56CX8;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DE 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Glucosyltransferase-T.
GN Name-gtf;
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1310;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=B13N;
RA Kuwahara N.;
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY966490; AAX76986.1; -; Genomic_DNA.
KW Transference.
SQ SEQUENCE 1506 AA; 167016 MW; 948882EF2C17C451 CRC64;

Query Match 48.8%; Score 3639; DB 2; Length 1506;
Best Local Similarity 49.3%; Pred. No. 6.4e-145;
Matches 724; Conservative 247; Mismatches 403; Indels 96; Gaps 26;

Qy 1 METKRRYKMKHKKHWVTAVAS-GLITL-GTTTLGSSVSA-----ETEQOTSQVVTQK 53
Db 1 MERKLHYLKHKKQWVTIAVASAGLASIVGAGLSQTVSVSADDLAKQAASQQAANAQ 60
Qy 54 SEDDKAASESSQTDAPKTKQ-----AQTEQTAQSQANVADTSTSTTKETPSQNI 103
Db 61 NEDEVASDAADTASAKATSEKEVVQSSDNTSETNQVETKQASAKESADAVAKAP----- 116
Qy 104 TQANSDDKTVNTYKSEEAOTSEERTKQSEEAOTASSQALTOAKABLTQKQRTAAQENK 163
Db 117 -----QAQPAATTSQ--VASSSESSVAPSKADKAAAGSVSQNEEEAA----- 156
Qy 164 NPVDLAAIPNVKOIDGKYVIGSDGQPKKFNALTNNKVLVFDKNTGALTDTTSOYQKQ 223
Db 157 -----LSLANIKKIDGKIYVMADGSKYKKNFALTVDGQMLYFDAKTGALSTSTSYSQ 211
Qy 224 LTKLNNDYTPHNOIVNFENTSLTIDNYVTADSWYRPKDILKNGKWTASSESLRPLLM 283
Db 212 LTPIVSDFSVNNKAFDSSEKSFELVDCYLTAESWYRPAKILENGKTWVDSKETDLRPVLM 271
Qy 284 SWPDKQTOIAYLNNYNNQGLGTGENYTDSSGESLNLAAOTVQVKIETKISQTOQOWL 343
Db 272 SWPNKQDTQVAYLNYMS-KALGKKEFTTETSQTLNTAAELIQAIEARVSKEQGKWL 330
Qy 344 RDIINSFVKTPNNWQSTESDTSAGEKDHLOGGALLYNSDKTAYANDYRLNRTTSQ 403
Db 331 REAAMAFVATQSRWNKDSQYDKA---DHLQGGALLYTNNNLTWANSNWRLLNRTTRQ 387
Qy 404 TGPKPTFEONSSGGYDFLLANDIDNSNPVVOAQLNWLHLYLMVYGSIVANDPEANFQVGR 463
Db 388 DGKTHYSKADKYGYEFELLANDVNSNPVVOAELNQHLYLMWGEIWMGDKANFQIR 447
Qy 464 VDAVDNNDLQIASDYLKAHYVDKSEKNAIHLISILEAWSDNDPQYNKDTKGAQLPI 523
Db 448 VDAVDNNDADTLQLYTNFYNSVYVGNKSEBAQALAHISVLEAWSYNDNDYNDQTNGAALAM 507
Qy 524 DNKLRLSLAYALTRPLEKASNKNEIRSGLEPVITN--SLNPRSAGKNSERMANIYFIR 581
Db 508 DNGLRFSLLYTLTRPI-----NERTPGMSTLIKSEYGLTDRTKNDKYGDTQPSYVYFVR 560
Qy 582 AHDSEVQTVIAKIIKAQINPKYTDGLTFLDELKQAFKIYNEDMRQAKCKYTQSGNIPTAYA 641
Db 561 AHDSEVQTVIAKIIKEKIDPTDGTFTLDQLKQAFEIYKMDNWSVKNKHYTHNIPTAYA 620
Qy 642 LMLSNDKSTIRLYYGDYSDQGYMATKSPYDAIDTLKARIKYAAGQDMKITVYEGD 701
Db 621 VMLSNNMESVTRVYVYGLDFTDDGQYMAKSPYDAINTLLRARIRYAAGQIMEHNSYKPS 680
Qy 702 -----KSHMD-----WDTYGLTVTSRYGTGANEATDQ--GSEATKTQGMNAVITSNPNSKLNQ 752
Db 681 AAMKAHPDAGNVLGNSEVLVSRFQDVMASDDMTGGKLAKTSGMFTLISNNPELELDV 740
Qy 753 NDKVIVNMGAAHKNQBYRPLLLTTKDLGTSYTSDDAAKSLYRKTNDKGLVDFASDIQ 812
Db 741 NEEIKVNVGKIHAQAYRPLLLTTDKGLOKYLNDSDTK-LTKIADKDGFTTFKGSSEIKGY 799
Qy 813 LNPQVSGYLAVWPVPGASNDQVRVAASNKANA-TQVYESSALDSQLIYEGFSNFQDF 871
Db 800 KQVEVNGYLSVWVPVPGAKADQIRVAPSTAAGKAKATYTAQALDSQLIYEGFSNFQDF 859
Qy 872 VTKDSYTNKIKIAQNVLPKSGWGTSEFMAPOVVSSEDSFLDSIIQNGYAFEDRDYLA 931
Db 860 VQKDSYTNKIKIAENTDLFKWGVTSFEMAPQVVSATDGFLLDSIIQNGYAFEDRDYLA 919
Qy 932 SKNNKYGSQODMINAVKALHSGIQVIADWVPDQIYNLPGKEVVVTRVNDYGEYRKDS 991

Db 920 SKNKKYGSREDLANALKALHAGIOAIADWPDQIYQLPGKEVVTASRVNDVGRVKVDOP 979
 QY 992 IKNTLYAANTKNGDYQAKYGGAFISELAAKYPSIFNRTOISNGKKIDPSEKITAWEKAK 1051
 Db 980 LVEKLYLANTKSGKDFQAKYGGEFIAELQKKYPFEMFTTKMISTGKTIDPSVKLKWSAK 1039
 QY 1052 YFNGTILGRGVYVLKDNASDKYFELKGNQTYLPKQMT-NKEASTGFVNDGNGMTPFYST 1110
 Db 1040 YFNGTIVLDRGTDYILSDDEGTGKYFTVNEKGDPLPASLTGKNDAKTGFYNDGKGIYVYTT 1099
 QY 1111 SGYQAKNSFVQDAKNWYFDDNNGHMVYQLQALNGEVQYFLSNGVOLRESFLENADGSKN 1170
 Db 1100 AGNKARSAPFTEA-GNTYFDYTGHMVTPGNVINTKPYFLPENGIMLKDAIKODEKGRSV 1158
 QY 1171 YFHLGNRYNSNGYSPDNDSKW-----RYFDASGVNAVGLKTINGNTQYFDQD 1218
 Db 1159 YGKTKGWYKGG-----RDNEWPMATDSKQMRFRHFRDYGPMISGLVITINQNVYYDEN 1213
 QY 1219 GYQVKAWITGSKKRYFDDGSGNMAVRPNKNDKGDWYLYNSDGLALVGQVQTINGKTY 1278
 Db 1214 GFQVKGEFVTDGOTRYFDQSSGNLVKGQFLN-KDGNWYLDQDQGLVAKGAQTIKGQKL 1272
 QY 1279 YFG-QDGKQIKGIITD-NGKLYFLANSSELARNIFATDSQNNWYFSGDGVAVTGSOT 1336
 Db 1273 YFDTKTVGVKGDVFDTKDGNTEFFYSGDGLDILGQFFSTGNMAWYADENGHVAKGAKT 1332
 QY 1337 IAGKCLYP-ASDGQVKVGSFVYNGKVHYHADSGELQVNRFEADKGN--WYLLDSNGE 1393
 Db 1333 IRGOKLVDTKTKQQAQKRFIRDKGVYYDADTGLVTNAFLETAKGSNQWYMGADGY 1392
 QY 1394 ALTGSRINDQRFVFTRE-GKQVKGDVAYD 1422
 Db 1393 AVKGNQTIQNHQWYFAETGQQAQKGIIVTD 1422

RESULT 12

Q8KZL5_9STRE PRELIMINARY; PRT; 1554 AA.
 ID Q8KZL5_9STRE PRELIMINARY; PRT; 1554 AA.
 AC Q8KZL5;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Glucosyltransferase.
 GN Name=gtfu;
 OS Streptococcus sobrinus.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1310;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RX MEDLINE=21958684; PubMed=11960691; DOI=10.1016/S0304-4165(01)00240-9;
 RA Hanada N., Fukushima K., Nomura Y., Senpuku H., Hayakawa M.,
 RA Mukasa H., Shiroza T., Abiko Y.
 RT "Cloning and nucleotide sequence analysis of the Streptococcus
 RT sobrinus gtfu gene that produces a highly branched water-soluble
 RT glucan.";
 RL Biochim. Biophys. Acta 1570:75-79(2002).
 RL EMBL; AB089438; BAC07265.1; -; Genomic_DNA.
 DR HSSP; P06653; 1H8G.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:0009250; P:glucan biosynthesis; IEA.
 DR InterPro; IPR002479; CW binding.
 DR InterPro; IPR003318; Glyco_hydro_70.
 DR Pfam; PF01473; CW binding_1; 5_
 DR Pfam; PF02324; Glyco_hydro_70; 1.
 KW Transferase.
 SQ SEQUENCE 1554 AA; 171676 MW; 6981BCC1DAE24A73 CRC64;
 Query Match 48.4%; Score 3615; DB 2; Length 1554;
 Best Local Similarity 49.9%; Pred. No. 6.8e-144;
 Matches 745; Conservative 217; Mismatches 418; Indels 112; Gaps 32;

QY 1 METKRYKMHKVKCHWTVAVAS-GLITL---GTTTLGSSVSAETEQQTSDKVVTKQSD 56
 Db 1 MEKKLHYKHVKHVKCHWTVIAVASIGLSVGVGCTVSAEDKVANDTTAQATGVGDTGQ--- 57
 QY 57 DKAASESQTDAPKT-----KQATQETQQAQS-QANVADTSTSIKTPSPQNITTOANSDD 111
 Db 58 DQATTNDANTNTDTDADQSANNTQDQAGSDOSNNQDQAKQDT-----ANTDR 106
 QY 112 KTVVTNKEEAQTSERTKQSEEAQTASSQALTOAKAELTKQRTAAQENKPNVDLAAI 171
 Db 107 NOADNSDNNQATDQATSPATDGTSVQRRDAANVATA-ADQEGQTASEQKSAAAL-SL 164
 QY 172 PNVKOIDGKYVYIGSDGQPKNFALTNNKVLVFKNTGALTDTSQYQFQKGLTKLNNDY 231
 Db 165 DNVKLIIDGKYVYVQADGSKKNFALTNGOMLYFSDTGALSTSTYSFSQGTNNLVDDP 224
 QY 232 TPNHNOI VNFENTSLTIDNYVTADSWYRPKDLKNGKWTWASSEDRLRPLLSMWPDKOT 291
 Db 225 SSHKAYDSTAKSPELVNGYLTANSWYRPAGILRNGQTWEASNENDLRPVLMSWPKDT 284
 QY 292 QIAYLNNNOQGLGTGENVTADSSQESLNLAQTVOVKLETKISQTOQTOWLRDIINSFV 351
 Db 285 QVAYVYNNKYLANSANETEVNTSETSQVDLNKEAQSIQTKIEQKITSDNSTQWLRTAMEAFV 344
 QY 352 KTOPNWNSTESDTSAGKDHLOGGALLYNSDKTAYANSDYRLNRTPTSTGPKPYE 411
 Db 345 AAQPKWNWSTE---NPNKGDHLOGGALLYNTNSDLTPWANSDYRLNRTPTQDGTKKYFT 401
 QY 412 DNSSGGYDFLLANDIDNSNPVVOAEOQLNWLHLYMNYGSIIVANDPEANFDGVRVADVNNY 471
 Db 402 EGGEGGYEFLNSDNDVNSNPVVOAEOQLNWLHLYMNGDIVMGDKANFDGVRVADVNNY 461
 QY 472 ADLQIASDYLAHYGVDKSEKNAVNHLSILEAWSDNDPOYNKDTKGALPDKLRLSL 531
 Db 462 ADLLQVYNYFKDNYKVTDSANALAHISILEAWSINDNQYNEGTALSDNSRLTS 521
 QY 532 LVALTR-PLEK-DASNKNEIRSLBPVITNSLN-NRSAGKNSERMANVIFIRAHDSVQ 588
 Db 522 LAVTKQPGQRIIDLSN-----LIGESVKNKERANDTAYGDTIPYTSFVRAHDSVQ 571
 QY 589 TVIAKIIIAQINPKTDGLTFTLDELKOAPKIYNEDMRQAKKYYTOSNIPTAYALMLSNKD 648
 Db 572 TVIAKIVKEKIDTNSDGYTFTLDQLKDAFKIYNEDMAKVNKTYTHYNI PAVALLSNME 631
 QY 649 SITRLYYGDMYSDGQYMATKSPYDAITLLKARIKYAAGQDMKITTYVEGDKSHMDWD 708
 Db 632 SVPRVYVYGLYTDGQYMAKSPYDAITATMLQGRVAYVSGQSEVHKVNGNG----- 686
 QY 709 YTGVLTSVRYGTGANEATD-QGSEATKTQGMVITSNPSLKNQNDKVINVMGAHKNQ 767
 Db 687 ----LSSVRYGQDLMSADDTQGTDLRSRTSGLVTLVSNDFNLDLG-GDSLTVNMGRAHQ 742
 QY 768 EYRPLLLTTKGLTSYTSDAAKSLYRKTNDKGLVDFDASDIQGYLNPQVSGYLAWVPV 827
 Db 743 AYRPLILGTGQGVQSVLKSDT-NIVKYTDANGNLFTTADDIKGYSTVDMSGYLAWVPV 801
 QY 828 GASDNQDVRVAASNKANATGVYESSALDSOLIIEGFSNFQDFVTKDSYTNKKIAQNV 887
 Db 802 GAKDQDVRVAADTNQKADGKSLKTSAAALDSQVIYEGFSNFQDFANNDADYTNKKIAENA 861
 QY 888 QLFKSGVTSFERMAPQYVSSSDGSLDSIIQNGYAFEDRYDLAMSNNKKNYGSODMINAV 947
 Db 862 DFFKGLGITSFENAPQYVSATDGSFLDSIIQNGYAFSDRYDLAMSNNKKNYGSKDDLAL 921
 QY 948 KALHKSIGQVIADWVPDQIYNLPGEVVTATRVNDYGEYRKDSEIKNTLYAANTKSGKD 1007
 Db 922 KALHANGIQAIADWVPDQIYQLPGEEVVTAKRTNSYGNPTFDAYINNALLYATNTKSGSD 981
 QY 1008 YQAKYGGAPFLSELAAYPSIFNRTOISNGKKIDPSEKITAWKAKYFNGTILGRGVYVL 1067
 Db 982 YQAKYGGAPFLSELAAYPSIFNRTOISNGKKIDPSEKITAWKAKYFNGTILGRGVYVL 1041
 QY 1068 KDNASDKYFELKGNQTYLPKQMT-NKEASTGFVNDGNGMTPFYSTGYSQAKNSFVQDAKGN 1126


```
Db 1042 SDDATGKYFTVNGDPLPASFTGDQNAKTFYDGTGMAYYSTSGNAVNSFIYEG-GH 1100
QY 1127 WYTFDNNHVMYGLQQL-NGEVOYFLSNGVOLRESFLENADGSKNYFGLGNRY-SNGYY 1184
Db 1101 YYYFDKDHVMTGSYKAEDGNDYYFLPNGIQMRDAIYQDAQNSYYVGRGILYKGDNWY 1160
QY 1185 SF-----DNDKWRYPDASGMVGLTKTINGNTQYFDQGYQVKGAWITGSDGKRYFDD 1239
Db 1161 PFVDPNNANKTVFRYPDANNMVAIGVNMVGTQYVFDENGFOAKGQLLTDKG-THYFDE 1219
QY 1240 GSGMVAVRFANDKNGDWYLLNSDGLALVGQVINGKTYFYG-ODGKQIKGLITD-NGK 1297
Db 1220 DNGAMAKNKFVN-VGDDWYIMDNGNAVKQYPPVNNQILYFNPETGVQVKGQFITDAQGR 1278
QY 1298 LKTYFLANSBELARNIFAT-----DSQNNWYFG-SDGVAVTGS 1334
Db 1279 TSYDANSALKSGSFTTPNGSDWYAENGVYVYKGFQVAENQDQYFDDTGTGKQAKGA 1338
QY 1335 QTITAGKLLYPASD-GKQVKGSTV-YNGKVHYHADSGELQVNRFEADKGNWYLDNSG 1392
Db 1339 AKVDGRDLFPNPSGVQVKGDFATDESNTSFYHGDNGKVKVGGFTTGNNAWYADNNG 1398
QY 1393 EALTGSQRI-----NDQVRF-TREGQVKGDDVAYD 1422
Db 1399 NLVKGQFQIDGKWHFDEVTGQAKGAALVNGQQLYFDVDSGIVQVKGDFVTD 1450

RESULT 13
GTF1_STRDO
ID GTF1_STRDO STANDARD; PRT; 1597 AA.
AC F11001;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
GN Name=gtf1;
OS Streptococcus downei (Streptococcus sobrinus).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1317;
RN [1]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RX STRAIN=MFE28;
RX MEDLINE=87308014; PubMed=3040686;
RA Ferretti J.J., Gilpin M.L., Russell R.R.B.;
RT "Nucleotide sequence of a Glucosyltransferase gene from Streptococcus
RT sobrinus MFE28.";
RL J. Bacteriol. 169:4271-4278(1987).
CC -!- FUNCTION: Production of extracellular glucans, that are thought to
CC play a key role in the development of the dental plaque because of
CC their ability to adhere to smooth surfaces and mediate the
CC aggregation of bacterial cells and food debris.
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-
CC fructose + (1,6-alpha-D-glucosyl) (n+1).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes
CC water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
CC forms of glucans.
CC -!- SIMILARITY: Belongs to the glucosyl hydrolase 70 family.
CC -!- SIMILARITY: Contains 19 cell wall binding repeats.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC
CC EMBL; M17391; AAC63063.1; -; Genomic DNA.
DR InterPro; IPR002479; Cell_wall_bd_put.
DR InterPro; IPR003318; Glyco_hydro_70.
```

```
DR Pfam; PF01473; CW binding 1; 4.
DR DR Pfam; PF02324; Glyco hydro 70; 1.
KW Dental caries; Glucosyltransferase; Repeat; Signal; Transferase.
FT SIGNAL 1 38 Potential.
FT CHAIN 39 1597 Glucosyltransferase-I.
FT REPEAT 1099 1132 A repeat.
FT REPEAT 1163 1213 AC repeat.
FT REPEAT 1227 1277 AC repeat.
FT REPEAT 1292 1342 AC repeat.
FT REPEAT 1352 1399 B repeat.
FT REPEAT 1406 1455 AC repeat.
FT REPEAT 1465 1512 B repeat.
FT REPEAT 1519 1568 AC repeat.
FT REPEAT 1582 1597 A repeat (incomplete).
FT REGION 39 1050 Catalytic (approximate).
FT REGION 1099 1597 1-25 A, 2 B and 5 AC repeats.
FT REGION 1099 1597 Glucan-binding (approximate).
SQ SEQUENCE 1597 AA; 177080 MW; B9E86A20086798E CRC64;

Query Match 48.1%; Score 3587.5; DB 1; Length 1597;
Best Local Similarity 48.5%; Pred. No. 1e-142;
Matches 733; Conservative 236; Mismatches 400; Indels 141; Gaps 31;

QY 1 METKRYKHKVKKHWTVAVASGLITLGTTLGSSV-SAETEQQTSDKVVTKSBDKA S9
Db 1 MEKNERFKMHKVKRWVTVSASA--TMLASALGASVASADTE-----TVSEDSNQA 50
QY 60 AESSOTDAPKTKQATEQTEQQAQSOANVADTSTSIKTPSQNIITQANSDDKTVNTKS 119
Db 51 VTADQT-----TTNQDTEQ-----SVAATATS-----EOSATAATDAQASDAQS 94
QY 120 EBAQTSEERTKQSEEAQTASS-----QALTQAKAELTKQRTAAQEN 162
Db 95 AAEQTQGTASTDTAAQTITANEAKWVTENENQVFTDEMLAEAKNVAT-----AESN 148
QY 163 KNPVLAALPNVKQIDGKYIYIGSGQPKKNFALTVNNKVLFPDKNTGALTDTFSQYF-K 221
Db 149 SIPDLAKMSNVKQVDGKYIYDQGNVKKNFVSVGEKIYFDE-TGAYKDTSKVEADK 207
QY 222 QG--LTKLNNDTYPHNQIVNPENTSLTIDNVYATDSWYRPKDLKNGKWTWASSDDL 279
Db 208 SGSDISKEETTFAANNRAYSTSAENFEADNLTADSWYRPKSIKDGKWTWESSKDD 267
QY 280 PLLMGWVPDQTOIAYLYNMYNQGLTGENYTADSSQESLNLAQTVQVKIETKISOTOQ 339
Db 268 PLLMAWVPDTEKRYVYVYVNVKV-VGIDKTYTAETSOADLTAAELVQARIEQKITTE 326
QY 340 TQWLRIINFSVYKTPNWNSTQESPTSGEKDHLGGALLYSN-SDKTAYANSDYRLNLR 398
Db 327 TKWLREALISAFVKTPQWNGSEKPYD---DHLQNGALKFDNQSDLTPTDQTSNYRLN 382
QY 399 TPTSQTG-----KPKYFEDNSSGGYDFLLANDIDNSNPVVOAQLNWLHLYMVGSI 454
Db 383 TPTNQTGSLDSRFTYNANDPLGGYELLANDVNDNSNPVVOAQLNWLHLYMVGSIYAK 442
QY 455 PEANFDGVRVDVAVNNADLLQIASDYKKAHYGVDSKKNANHLISILRAWSNDNDPQY 514
Db 443 ADANFDSIRVDVAVNNADLLQISDDYLKAAVGIKNNKNNVNNHVSIVAWSDNDTPY 502
QY 515 DTGKAQLPIDNKLRLSLLYALTRPLEKQASKNKNEIRSGLEPVITNSLNRSRSGKNS 574
Db 503 DGDGDLNMMNDKFRLSMLWSLAKPLDK-----RSLGNPLIHNSLVDRVDDREVE 554
QY 575 ANYIPIRAHDSVQVVIKIIKAOINPKYDGLTFTLDELKQAFKYNEDMRQAKKYTOS 634
Db 555 PSYSFARHDSVQVQILIRDIKAEINPNAFGYSFTQDEIDQAFKYNEDLKTKKYTH 614
QY 635 NIPTAYALMLNKSQITRLYVGDWYSDGQVWATKSPYYDAITDILLKARIKYAAGQDN 694
Db 615 NVPLSYTLLLTNKGSIPIRVYIGDMFTDQGYNANKVTNYDAIESLLKARKMKVAGQ 674
QY 695 ITYVEGDKSHMDWDTGVLTYSVRYGTGANEATDQGEATKTQGMVITNSNPNLSKLN 754
```



```
Db 675 -NYQIGNGE-----ILTSVRYGKALKQSKDQATRTTSVGVVGMGNQNFSLD-GK 724
QY 755 KIVVNMGAHQEQEYRPLLLTKDGLTSYTSDAAK--SLYRKTNDKGBLVDFASDIQY 812
Db 725 VVALANMGAHQEQEYRALMVSTKDGVTATDADASKAGLVKRTDENGSLYFLNDLKG 784
QY 813 LNPQVSGYLAVVPVPGASDNQDVRVAASKANATQVYESSALDSQLIYEGFSNQDFV 872
Db 785 ANPQVSGFLQVVPVPGAAQDDQIRVAASDTADGKSLHQDAMDSRVNMFEGFSNQFSA 844
QY 873 TKDSYTNKIKAQNVQLFKSWGVTSPMAPQVYSSSDGSLDSIIQNGYAFEDRDVLAMS 932
Db 845 TKEEETVNVIANNVDFVSWGITDFEMAPQVYSSDTGQFLDSVLQNGYAFTRDYLGS 904
QY 933 KXNKYGSQDDMINAVKALHKSGIQVIADWPQIYNLPKEVVTATRVNDYGEYRKDSI 992
Db 905 KANKYGTADQLKAIKALHAKGLKVNADWPQMYTFPKQEVVTVTRTDKFKPIAGSOI 964
QY 993 KNTLYAANTKSKGDKYQAKYGGAFSLSELAAKYPSIFNRTOISNGKKIDPSEKITAWKAY 1052
Db 965 NNSLYVTDTKSGDDYQAKYGGAFSLDELKEKYPELFTFKQISGQADPSVKIKQWSAKY 1024
QY 1053 FNGTNIILRGVGVVLKDNASDKYFELKGNQTYLPKQMTNKEASTGPNVNDGNGMTF-YSTS 1111
Db 1025 FNGSNILRGADYVLSQASNKYLVNDDKLFPLKTLQGVVESGIRFDGTGVVNSSTI 1084
QY 1112 GYQAKNSFVQDAGKNWYFDNNHMYGLQOLNGEVOYFLSNGVQLRBSFLENADGSKNY 1171
Db 1085 GEKVTSDFITEA-GNLYYFGQDGYMTGAQNIKGSNYFLANGAALRNTVYTDQAGQNH 1143
QY 1172 FGHGNGRYNGYYSFNDKSKWYFDASGYMAVGLKTINGTOYEDODGYOVKGAWITGSD 1231
Db 1144 YGNDGKRYENGVOQFGNDS-WRYF-KNGVMALGLTTVDGHVQVFDKGQVQAKDIIVTRD 1201
QY 1232 GKRYFDDGSGNMAVNRFANDKNGDWYLYNSDGIALVGVQTINGKTYFQDQKQIKGKI 1291
Db 1202 GKRYFDDQNGNAVTTVFADKTGHWYLYLKGQVAVTGAQTVKGQHLYPEANGQVKGDF 1261
QY 1292 IT-DNGKLYFLANSSELARNIPATDSQNNWYFSGDGVAVTGSQTIAGKLYFASDGKQ 1350
Db 1262 VTAKDGKLYFDYDSDGMWNTTFIEDKAGNWFYLGKDGAAVTGAQTIKGQKLYFKANGQ 1321
QY 1351 VKGSFV-TYNGKVHYHADSGELQVNR-----FEAD-----KQGN- 1384
Db 1322 VKGDIVKADAGKIRYIDAQTEGQVFNKSVSVNGKTYFSGSDGTGAQTPANKGQTFKQDGS 1381
QY 1385 -----WYILDSNGEALTGSRINDQRVFFTRGKQVKGD-V 1419
Db 1382 VLRFYNLEGQYVSGSGWYETAHEWYVKS-GKVLGTGAQTIGNQRYVFXDNGHQVKGQLV 1440
QY 1420 AYDERLLVY 1429
Db 1441 TGNDGKLRVY 1450
```

RESULT 14

```
Q59983 9STRE
ID Q59983.9STRE PRELIMINARY; PRT; 1590 AA.
AC Q59983;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5).
GN Name=gtfii;
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1310;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=QM2176;
RX MEDLINE=94146405; PubMed=8312602;
RA Sato S., Inoue M., Handa N., Aizawa Y., Isobe Y., Katayama T.;
```

"DNA sequence of the glucosyltransferase gene of serotype d Streptococcus sobrinus";
DNA Seq. 4:19-27(1993).
[2]

PROTEIN SEQUENCE.

RA Moore G., Hefta S.A., Paxton R.J., Shively J.E., Lee T.D.;
MEDLINE=91224988; PubMed=1827439;
"Isolation and sequence of an active-site peptide containing a catalytic aspartic acid from two Streptococcus sobrinus alpha-glucosyltransferases";
J. Biol. Chem. 266:8916-8922(1991).
RL EMBL; D13858; BAA02976.1; -; Genomic_DNA.
DR PIR; A39841; A39841.
DR HSSP; P06653; 1HCX.
DR GO; GO:0047849; F:dextranucrase activity; IEA.
DR GO; GO:0016757; F:transferase activity, transferring glucosyl. .; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1_2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Glucosyltransferase; Signal; Transferase.
FT SIGNAL 1 38 Potential.
FT CHAIN 39 1590 Glucosyltransferase-I.
SQ SEQUENCE 1590 AA; 175956 MW; C3C83A57CF3C2B0E CRC64;

Query Match 47.9%; Score 3575; DB 2; Length 1590;
Best Local Similarity 48.4%; Pred. No. 3.4e-142;
Matches 719; Conservative 240; Mismatches 408; Indels 118; Gaps 26;

```
QY 1 METKERYKMHVKKHWTVAVASGLITLGTTLGSSVSAAETEQTSDKVVTQKSEDDKAA 60
Db 1 MEKNVRFKMKHVKKHWTVLSVASA--TMLASALGASVAS-----ADTDITAS 44
QY 61 SESSQTDAPKTKQAQTEQTQAOSQANVADTSTSIKTETPSQNTITQANSDDKTVNTKSE 120
Db 45 DDSNQA---VVTGDTQTNNAQTDQTSIAATATS-EQSASTDAATDQASAAEQTQGTAST 100
QY 121 EAQTSERTKQSEEAQ-----TTASSQALTAQKAEITKQRTAAQENKNPVDLAAIPVKQI 177
Db 101 D--TAAQTITNANEAKWPTENENQGTDEMLAEAKNVAT-AESDIPSLDLAKMSNVKQV 157
QY 178 DGKYYYIGSDGQPKNFALTNNKLVLPDKNTGALTDTTSQYQFKGLTTLKNDND---YTPH 234
Db 158 DGKYYYDQDGNVKNFVSGVKIYYFDE-TGAYKDTSKVDADKSSSAVSQATIFAAN 216
QY 235 NQIVNFENTSLEITDNYVTADSWYRPFKQILKNGKTTWTASSEDRLRPLLSMWPDKQQTIA 294
Db 217 NRAYSTSAKNFEAVDNYLTADSWYRPKSILKDGKTWTESGKDDFRPLLMWMPDPTETKRN 276
QY 295 YLYNNQOGLGTGENYTADSSQESINLAAQTVOVKIETKISQTOQTOWLRDINSFVKTKQ 354
Db 277 YVYNNKVV-VGIDKTYTAETSOADLTAAAEVLQARIEQKITSENNTKWLREASIAFVKTKQ 335
QY 355 PNWNSQTSQTSAGEKQHLQGGALLYSN-SDKTAVANSQYRLNLRTPETSTQGT---KPKY 409
Db 336 PWNQGESEKPYD-----DHLQNGALLFDNQTDLTPTDQSNRYRLNLRTPETSTQGLSDRFFY 391
QY 410 FEDNSGGYDFLLANDIDNSFPVQAEQLNWLHLYLMNYGSIIVANDPEANFDGVRVDAVDN 469
Db 392 NFNDPLGGYDFLLANDVDNSFPVQAEQLNWLHLYLNFGSIIVANDADANFDSIRVDAVDN 451
QY 470 VNADLLQIASDYLKAHYGVKSEKNAIHLHSILEAWSNDNDPQYNKDTKAQLPIDNKLRL 529
Db 452 VDADLLQIASSDYLKAAAYGIDKNNKNNHVSIVEAWSNDNTPYLHDDGDLNMMNMFRL 511
QY 530 SILLYALTRPLEKDASKNKEIRSGLEPVTINSLNNSAEGKNSERMANYIFIRAHDSVQV 589
Db 512 SMLWSLAKPLDK-----RSLGPLIHNSLVDRVDDREVETVPSYSFARAHDSVQD 563
QY 590 VIAKIKAIQINPKTDGLTFLDELKQAFKIYNNEDMRQAKKYQTSNIPYAYLMLSNKDS 649
Db 564 IIRDIKAEINPNSFGYSPTQBEIEQAFKIYNNEDLKKTKDKYTHYNVPLSYTILLTNKGS 623
```


Db 624 IPRVYDMFTDDGQYMANVTNYDAIESLLKARMKYVSGQAMQ-NYQIGNGE----- 676

Qy 710 TGVLTSVRYGTGANRATDQGSBATKTQGMVITSSNNPSLKLNONDKVIVNMGAAHKNQY 769

Db 677 --ILTSVRYGKALKQSDKGDATRTTSGVGVVGMQPNFSLD-GKVVALNMGAAHANQY 733

Qy 770 RPLLLTTKDLTSYTSDDAAK--SLYRKTNDEKELVFDASDIQGYLNPQVSGYLAVWV 827

Db 734 RALMVSTKDGVAITYATADASKAGLVKRTDENGILYFLNDDLKGVANPQVSGFLQVWV 793

Qy 828 GASDNQDVRVAASNKANATQVYESSALDSOLIYEGFSNFQDFVTKDSDYTNKKIAONV 887

Db 794 GAADDQDINVAASDTASTDGLSHQDAANDSRVMEGFSNFQSFATKEEYTNVVIANNV 853

Qy 888 QLFKSWGVTSEFMAPOVVSSEDSFLDSIIQNGYAFEDRYDLAMSKNNKYGSQQDMINAV 947

Db 854 DKFVSWGITDFEMAPQVVSSTDQFLDSVIQNGYAFEDRYDLGMSKANKYGTADQLVKAI 913

Qy 948 KALHKSIOVIADWVDDQIYNLPKEVVTATRVNDYGEYRKDSEIKNTLYAANTKSKNGK 1007

Db 914 KALHAKGLKVMADWVDDQYTPKQEVVTVTRDKFQPIAGSQINHSLYLVTDTKSSGDD 973

Qy 1008 YQAKYGGAFLSELAAKYPSIFNRTOISNGKKIDPSEKITAWKAKYFNGTNILGRGVGYVL 1067

Db 974 YQAKYGGAFLDELKEXYPELFTYKQISTQADPSVKIKQWSAKYFNGSNILGRGADYVL 1033

Qy 1068 KDNASDKYFELKGNQTYLPKQMTNKEASTGFVNDGNGMTFYST-SCYQAKNSFVQDAKGN 1126

Db 1034 SDQVSNKYFNVASDRTLFLPSSLLGKVESGIRYDGKGIYNSSATGDQVKASFITEA-GN 1092

Qy 1127 WYFFDNNGHMVYGLQOLNGEVQYFLSNGVQLRESFLENADGSKNYFGHLGNRYSNGYYSF 1186

Db 1093 LYFFGKDGMYMTGAQTINGANYFFLENGTALRNTIYTDQGNSHYANDGKRYENGYYQF 1152

Qy 1187 DNDSKWRYFDASGVMAVGLKTINGNTQYFDQDGYQVKGAWITGSDGKKRYFDDGSGNMAV 1246

Db 1153 GND--WRYF-KDGNMAVGLTIVDGNVQYFDKGQVQAKDKIIVTRDGKRYFDOHNGNAV 1209

Qy 1247 NRPANDKNGDWYLLNSDGLVGVQTINGKTYFFQDQKQIKGKIITDN-GKLKYFLANS 1305

Db 1210 NTFIADKTGHWTYLGKDGVAVTGAQTVGKQLYFEANGQVKGDFVTSHEGKLYFYDVDS 1269

Qy 1306 GELARNIPATDSQNNWYFFSGDGAVTGSQTTAGKLYFASDGKQVKGFSVT-YNGKVHY 1364

Db 1270 GDMWTDFTIEDKAGNWFYLGKGAAVSGAQTIRGQKLYPKAYQQQYKGDIVRGTDGKIRY 1329

Qy 1365 YHADSGELQVNRFEADKQ----- 1383

Db 1330 YDAKSGEVPNKTVKAAADSKTYVIGNGVAVDPSVVKQTFPKDASGALRFPYNLKGQLVTG 1389

Qy 1384 -----NWYYLDSNGEALTSQSRINDQRFVFTREGKQVKGDV 1419

Db 1390 SGWYETANHDWVYIQS-GKALTGEQTINGQHLFYFKEDGHQVKQL 1433

Search completed: February 11, 2006, 19:38:20
Job time : 199.956 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 18:56:56 ; Search time 158.264 Seconds
(without alignments)
4414.224 Million cell updates/sec

Title: US-10-797-821-37

Perfect score: 8349

Sequence: 1 MEKNVRFKMHVKRWVTLT.....VYQYYFYFGDGIYRGWN 1590

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 segs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*

- 1: Geneseq1980s.*
- 2: Geneseq1990s.*
- 3: Geneseq2000s.*
- 4: Geneseq2001s.*
- 5: Geneseq2002s.*
- 6: Geneseq2003as.*
- 7: Geneseq2003bs.*
- 8: Geneseq2004s.*
- 9: Geneseq2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8349	100.0	1590	7	Add93657 Streptoco
2	8349	100.0	1590	9	Adx37280 Streptoco
3	8256	98.9	1592	2	Aar32925 Glucosylc
4	4672.5	56.0	1476	5	Aau79284 Streptoco
5	4580	54.9	1475	5	Aau98027 S. mutans
6	4580	54.9	1475	7	Add93654 Streptoco
7	4580	54.9	1475	9	Adx37277 Streptoco
8	4579	54.8	1475	5	Aau98030 S. mutans
9	4576	54.8	1475	5	Aau98040 S. mutans
10	4575	54.8	1475	5	Aau98031 S. mutans
11	4574	54.8	1475	5	Aau98033 S. mutans
12	4573	54.8	1475	5	Aau98032 S. mutans
13	4568	54.7	1475	5	Aau98035 S. mutans
14	4568	54.7	1475	5	Aau98034 S. mutans
15	4566	54.7	1475	5	Aau98036 S. mutans
16	4560	54.6	1475	5	Aau98037 S. mutans
17	4553	54.5	1475	5	Aau98039 S. mutans
18	4550	54.5	1475	5	Aau98038 S. mutans
19	4214.5	50.5	1375	5	Aau98028 S. mutans
20	4214.5	50.5	1375	5	Aau79288 Streptoco
21	4214.5	50.5	1375	7	Add93655 Streptoco
22	4214.5	50.5	1375	9	Adx37278 Streptoco
23	3587	43.0	1430	5	Aau98044 S. mutans
24	3583	42.9	1430	5	Aau98045 S. mutans

25	3582	42.9	1430	5	AAU98041	S. mutans
26	3580	42.9	1430	5	AAU98043	S. mutans
27	3578	42.9	1430	5	AAU98042	S. mutans
28	3575	42.8	1430	5	AAU98029	S. mutans
29	3575	42.8	1430	7	ADD93656	Streptoco
30	3575	42.8	1430	9	ADX37279	Streptoco
31	3525	42.2	1499	7	ADC54806	Protein S
32	3502.5	42.0	1554	7	ADD93658	Streptoco
33	3502.5	42.0	1554	9	ADX37281	Streptoco
34	3451.5	41.3	1017	5	AAU79285	Streptoco
35	3325.5	39.8	1518	7	ADD93660	Streptoco
36	3325.5	39.8	1518	9	ADX37283	Streptoco
37	3282.5	39.3	2835	5	ABB98574	Dextran 8
38	3282.5	39.3	2835	6	ABR55594	Amino aci
39	3245	38.9	1577	2	AAR91047	Alpha-D-g
40	3087.5	37.0	1527	5	AAU80055	Leuconost
41	3085.5	37.0	1527	7	ADC54807	Leuconost
42	3053	36.6	1477	9	ADY72696	Mutant de
43	3052	36.6	1477	9	ADY72733	Mutant de
44	3049	36.5	1477	9	ADY72732	Mutant de
45	3031	36.3	1365	7	ADD93659	Streptoco

ALIGNMENTS

RESULT 1

ADD93657

ID ADD93657 standard; protein; 1590 AA.

XX AC ADD93657;

XX DT 29-JAN-2004 (first entry)

XX DE Streptococcus sobrinus glucosyltransferase-I.

XX KW Glucosyltransferase; enzyme; vaccine; antitoxin; epitope; immunogen.

XX OS Streptococcus sobrinus.

XX PN WO2003075845-A2.

XX PD 18-SEP-2003.

XX PF 07-MAR-2003; 2003WO-US006962.

XX PR 07-MAR-2002; 2002US-0363209P.

XX PR 08-AUG-2002; 2002US-0402483P.

XX (FORS-) FORSYTH INST.

XX PI Smith DJ, Taubman MA;

XX DR WPI; 2003-845091/78.

XX PT Composition useful as vaccines for dental caries comprises a fragment of a glucan binding protein-B binding to a major histocompatibility complex class II protein.

XX Claim 16; Page 14; 49pp; English.

XX CC The present sequence is the protein sequence of Streptococcus sobrinus glucosyltransferase-I (GTF-I). Peptide fragments of GTF-I, especially from the catalytic domain of the polypeptide, can be used in immunogenic compositions and subunit vaccines for dental caries. These compositions comprise a major histocompatibility complex (MHC) class II protein-binding peptide from S. mutans glucan binding protein-B (GbpB) covalently linked with a peptide fragment of a streptococcal glucosyltransferase. The compositions are used in a claimed method of eliciting production of an antibody in a mammal. Dieptopic or multieptopic polypeptides can be prepared synthetically or by recombinant DNA technology. Antibodies raised against MHC class II binding fragments of GbpB can be used in passive immunisation.

XX Smith DJ, Taubman MA;
 XX WPI; 2005-151644/16.
 XX
 XX New composition comprising a fragment of a glucan binding protein-B
 PT (GbpB) that binds to MHC class II protein, and a biocompatible
 PT microparticle, useful for producing an antibody (claimed) for immunizing
 PT mammals against dental caries.
 XX
 XX Claim 7; SEQ ID NO 37; 73pp; English.
 XX
 CC The invention relates to a composition comprising a fragment of a glucan
 CC binding protein-B (GbpB) and a biocompatible microparticle, where the
 CC fragment binds to a major histocompatibility complex (MHC) class II
 CC protein. The composition is useful for producing an antibody for
 CC immunizing mammals against dental caries. This sequence corresponds to a
 CC Streptococcus sobrinus GbpB protein of the invention.
 XX
 SQ Sequence 1590 AA;
 Query Match 100.0%; Score 8349; DB 9; Length 1590;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1590; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MEKNVRFKHKVKKRWVTLVSASATMLASALGASVASADTDTASDSSNQAVVTGDDQTTNN 60
 DB 1 MEKNVRFKHKVKKRWVTLVSASATMLASALGASVASADTDTASDSSNQAVVTGDDQTTNN 60
 QY 61 QATDQTSIAATATSEOSASTDAATDQASAAEQGTASTDTAAQTITNANEAKWPTEN 120
 DB 61 QATDQTSIAATATSEOSASTDAATDQASAAEQGTASTDTAAQTITNANEAKWPTEN 120
 QY 121 ENQGFTEMLAEAKNVATAESDSIPSDLAKMNVKQVDGKYYYDODGNVKNFVSVGD 180
 DB 121 ENQGFTEMLAEAKNVATAESDSIPSDLAKMNVKQVDGKYYYDODGNVKNFVSVGD 180
 QY 181 KIYFDETGA YKDTSKVDADKSSAVSQNATIPANNRAYSTSAKNFEAVDNYLTADSWY 240
 DB 181 KIYFDETGA YKDTSKVDADKSSAVSQNATIPANNRAYSTSAKNFEAVDNYLTADSWY 240
 QY 241 RPKSILKDGKWTESGKDFRPLLMAWPDTETKRNYYVMNKNVGVGIDKTYTAETSQADL 300
 DB 241 RPKSILKDGKWTESGKDFRPLLMAWPDTETKRNYYVMNKNVGVGIDKTYTAETSQADL 300
 QY 301 TAAAEVLQARIIEQKITSENNTKWLRSAISAFVKTPQPMNGESEKPYDDHLQNGALLFDNQ 360
 DB 301 TAAAEVLQARIIEQKITSENNTKWLRSAISAFVKTPQPMNGESEKPYDDHLQNGALLFDNQ 360
 QY 361 TDLTPTQSNYRLNRPNTNQTGSLDSRFTYNPNDPLGGYDFLLANDVDNSNPVQAEQL 420
 DB 361 TDLTPTQSNYRLNRPNTNQTGSLDSRFTYNPNDPLGGYDFLLANDVDNSNPVQAEQL 420
 QY 421 NWLYLLNFGSYYANDADANFDSIRVDVNDVDADLLQISSDYLLKAAAYGIDKNNKNNH 480
 DB 421 NWLYLLNFGSYYANDADANFDSIRVDVNDVDADLLQISSDYLLKAAAYGIDKNNKNNH 480
 QY 481 VSIWEAWSNDPTPYLHDDGDNLMNDKFLSLMLSLAKPLDKRSLNPLIHNSLVDRV 540
 DB 481 VSIWEAWSNDPTPYLHDDGDNLMNDKFLSLMLSLAKPLDKRSLNPLIHNSLVDRV 540
 QY 541 DDRETVTPSYSFARAHSEVDQIIRDIIEKAEINPNSFGYSFTQEBIEQAFKIYNEDLKK 600
 DB 541 DDRETVTPSYSFARAHSEVDQIIRDIIEKAEINPNSFGYSFTQEBIEQAFKIYNEDLKK 600
 QY 601 TDKKTYHNVPLSYTLNLLTNKGSIPRVYVYGMFTDDGQYMANKTVMNYDAIESLLKARMKY 660
 DB 601 TDKKTYHNVPLSYTLNLLTNKGSIPRVYVYGMFTDDGQYMANKTVMNYDAIESLLKARMKY 660
 QY 661 VSGGQAMQNYQIENGCEILTSVRYGKALKQSDKGDATTTSGVGVVMGNQPNFSLDGKVV 720
 DB 661 VSGGQAMQNYQIENGCEILTSVRYGKALKQSDKGDATTTSGVGVVMGNQPNFSLDGKVV 720

QY 721 ALANGAHAHQEYRALMVSTKDG VATYATDADASKAGLVKRTDENG YLYFLNDDLKGVAN 780
 DB 721 ALANGAHAHQEYRALMVSTKDG VATYATDADASKAGLVKRTDENG YLYFLNDDLKGVAN 780
 QY 781 PQVSGFLOVWPVGAADDQDIRVAASDTASTDGSLHODAAMSRVNMFEGSFNQSFATK 840
 DB 781 PQVSGFLOVWPVGAADDQDIRVAASDTASTDGSLHODAAMSRVNMFEGSFNQSFATK 840
 QY 841 EEEYTNVVIANNVDKPVSWGITDFEMAPQVSVSTSDGQFLDSVIONGVAFTRDYDLGMSKA 900
 DB 841 EEEYTNVVIANNVDKPVSWGITDFEMAPQVSVSTSDGQFLDSVIONGVAFTRDYDLGMSKA 900
 QY 901 NKYGADQLVKALIKALHAKGLKVMADWVPDQMTYFPKQEVVTVTRTDKFGKPIAGSQINH 960
 DB 901 NKYGADQLVKALIKALHAKGLKVMADWVPDQMTYFPKQEVVTVTRTDKFGKPIAGSQINH 960
 QY 961 SLVYVTTKSGDDYQAKYGAGFLDELKEKYPELFTKKQISTGOAIDPSVKIKOWSAKYFN 1020
 DB 961 SLVYVTTKSGDDYQAKYGAGFLDELKEKYPELFTKKQISTGOAIDPSVKIKOWSAKYFN 1020
 QY 1021 GSNILGRGADYVLSDOVSNKYFNVASDTLFLPSSLLGKVVEGIRYDGKVIYNSSATGD 1080
 DB 1021 GSNILGRGADYVLSDOVSNKYFNVASDTLFLPSSLLGKVVEGIRYDGKVIYNSSATGD 1080
 QY 1081 QVKASPIITEAGNLVYFGKDGVMVTGAQTINGANYFFLENGTALRNTIYTTDAQNSHYAN 1140
 DB 1081 QVKASPIITEAGNLVYFGKDGVMVTGAQTINGANYFFLENGTALRNTIYTTDAQNSHYAN 1140
 QY 1141 DGKRYENGYQQFGNDWRYPKOGNNAVGLTTVDGNVQYFDKDGVOAKDKIIVTRDGKVRYP 1200
 DB 1141 DGKRYENGYQQFGNDWRYPKOGNNAVGLTTVDGNVQYFDKDGVOAKDKIIVTRDGKVRYP 1200
 QY 1201 DQHGNGNAATNTFADTKTGHWYLYGKDGVAVTGAQTVGKOKLYPEANGQQVKGDFVTSDEG 1260
 DB 1201 DQHGNGNAATNTFADTKTGHWYLYGKDGVAVTGAQTVGKOKLYPEANGQQVKGDFVTSDEG 1260
 QY 1261 KLYFYDVDSGDMWTDFTIEDKAGNWFYLGKDGAAVTGAQTIHQKLYFKANGQQVKGDI 1320
 DB 1261 KLYFYDVDSGDMWTDFTIEDKAGNWFYLGKDGAAVTGAQTIHQKLYFKANGQQVKGDI 1320
 QY 1321 KGTGDKIRYDAKSGEQVFNKTVKAADGKTYVIGNDGVAVDPSVVGQTFKDGASGALRFY 1380
 DB 1321 KGTGDKIRYDAKSGEQVFNKTVKAADGKTYVIGNDGVAVDPSVVGQTFKDGASGALRFY 1380
 QY 1381 NLKGQLVTCGSGWYETANHDWVYIQSGKALTGEOTINGQHLVYFKEDGHQVKGQLVTGTDGK 1440
 DB 1381 NLKGQLVTCGSGWYETANHDWVYIQSGKALTGEOTINGQHLVYFKEDGHQVKGQLVTGTDGK 1440
 QY 1441 VRYDANSGDQAFNKSVTYNGKTYTFGNDGTAQTAGNPKGQTFKOGSDIRFYSMEGQLVT 1500
 DB 1441 VRYDANSGDQAFNKSVTYNGKTYTFGNDGTAQTAGNPKGQTFKOGSDIRFYSMEGQLVT 1500
 QY 1501 GSGWYENAGQMLYVKNQKVLTLGLTVGSRQVYFDENGIOAKGKAVRTSDGKIRYFDENS 1560
 DB 1501 GSGWYENAGQMLYVKNQKVLTLGLTVGSRQVYFDENGIOAKGKAVRTSDGKIRYFDENS 1560
 QY 1561 GSWITNQWKFVYQYVYFGNDGARIYRGWN 1590
 DB 1561 GSWITNQWKFVYQYVYFGNDGARIYRGWN 1590
 RESULT 3
 AAR32925
 ID AAR32925 standard; protein; 1592 AA.
 XX
 AC AAR32925;
 XX
 DT 28-JUN-1993 (first entry)
 XX
 DE Glucosyltransferase I.
 XX
 KW GT-1; Streptococcus; dental; caries.
 XX

OS Streptococcus sobrinus.
XX JP05023188-A.
PN XX
XX XX
PD 02-FEB-1993.
XX XX
XX 25-JUL-1991; 91JP-00186592.
XX XX
XX 25-JUL-1991; 91JP-00186592.
XX XX
XX (KATO/) KATO K.
XX (FUKU/) FUKU I.
XX WPI; 1993-079449/10.
XX N-PSDB; AAQ37760.
XX
XX DNA sequence glucosyl transferase-I - comprises Streptococcus sobrinus
XX DNA sequence with at least one nucleotide added or deleted.
XX
XX Claim 13; Page 15; 29pp; Japanese.
XX
XX The DNA sequence from Streptococcus sobrinus strain 6715 encodes
XX glucosyl transferase-I (and mutants). The DNA was obtd. by treating S.
XX sobrinus 6715 with mutanolysin, extracting the chromosomal DNA, partially
XX digesting with SauAI and fractionating on agarose gel. The 3-5 kbp
XX fragment was ligated into pUC18 and E. coli JM109 transformed with it. A
XX GT-1 expressing clone was isolated and sequenced. The clone may be used
XX in the development of a drug for dental caries
XX
XX
XX
SQ Sequence 1592 AA;

Query Match 98.9%; Score 8256; DB 2; Length 1592;
Best Local Similarity 98.9%; Pred. No. 0;
Matches 1574; Conservative 6; Mismatches 10; Indels 2; Gaps 1;

QY 1 MEKNVFPKHKVKKRWVTLVSASATMLASALGASVASADTDASDDSNQAVTGDQTTNN 60
DB 1 MEKNVFPKHKVKKRWVTLVSASATMLASALGASVASADTDASDDSNQAVTGDQTTNN 60

QY 61 QATDQTSIAATATSEOSASTDAATDQASAAEOTQGTASTDTAAQTITTANAEAKWVPTEN 120
DB 61 QATDQTSIAATATSEOSASTDAATDQASAAEOTQGTASTDTAAQTITTANAEAKWVPTEN 120

QY 121 ENQGFTDEMLAEAKNVAATESDPSDLAKMSNVKQVDGKYYYDQDGNVKNFAVSVDG 180
DB 121 ENQGFTDEMLAEAKNVAATESDPSDLAKMSNVKQVDGKYYYDQDGNVKNFAVSVDG 180

QY 181 KIYYPDETGAUKDTSKVADAKSSAVSONATIPAAANNRAYSTSAKNFEAVDNYLTADSWY 240
DB 181 KIYYPDETGAUKDTSKVADAKSSAVSONATIPAAANNRAYSTSAKNFEAVDNYLTADSWY 240

QY 241 RPKSILKDGKWTWESGKDFRPLLMAWPDTETKRNVMNKNVGVGIDKTYTAETSQADL 300
DB 241 RPKSILKDGKWTWESGKDFRPLLMAWPDTETKRNVMNKNVGVGIDKTYTAETSQADL 300

QY 301 TAAAEIVQARIEQKITSENNTKWLREAI SAFVKTQPOWNGESEKPYDDHLQNGALLFDNQ 360
DB 301 TAAAEIVQARIEQKITSENNTKWLREAI SAFVKTQPOWNGESEKPYDDHLQNGALLFDNQ 360

QY 361 TDLTPDTQSNRYLLNTRPTNQTGSLDSRFTYNPNDFLGGDYFLANDVDNSNPVQAEOL 420
DB 361 TDLTPDTQSNRYLLNTRPTNQTGSLDSRFTYNPNDFLGGDYFLANDVDNSNPVQAEOL 420

QY 421 NWLHYLLNFGSIIYANDADANFDSIRVDADVNDADLLQISSDYLKAAYGIDKNNKANHH 480
DB 421 NWLHYLLNFGSIIYANDADANFDSIRVDADENVDADQI QISSDYLKAAYGIDKNNKANHH 480

QY 481 VSIVEAWSNDPTPYLHDDGDNLMNDKNPRLSMLSLAKPLDKRSGLNPLIHNLSVDREV 540
DB 481 VSIVEAWSNDPTPYLHDDGDNLMNDKNPRLSMLSLAKPLDKRSGLNPLIHNLSVDREV 540

QY 541 DDREVEVPSYSPARAHSEVQDIIRDIIKAEINPNPSFGYSEFTQBEIEQAFKIYNEDLKK 600
DB 541 DDREVEVPSYSPARAHSEVQDIIRDIIKAEINPNPSFGYSEFTQBEIEQAFKIYNEDLKK 600

AAU79284 standard; protein; 1476 AA.
AAU79284;
13-AUG-2002 (first entry)
Streptococcus mutans monoclonal antibody-related protein #1.
XX Antibody; dental caries; water insoluble glucan synthetase; anti-carries;
KW glucosyl transferase-B; immunotherapy.
XX Streptococcus mutans.
OS Streptococcus mutans.
XX JF2002114709-A.
XX 16-APR-2002.
XX 04-OCT-2000; 2000JP-00304889.
XX 04-OCT-2000; 2000JP-00304889.
PR (UYN1-) UNIV NIPPON.
XX WPI; 2002-448885/48.
XX Anti-carries agent composed of a monoclonal antibody against an inhibitory
PT enzyme against water insoluble glucan synthetase of glucosyl transferase-
PT B (GTF-B) of Streptococcus mutans.
XX Claim 3; Page 13-16; 28pp; Japanese.
XX The invention relates to a monoclonal antibody against dental caries and
CC an anti-carries agent composed of a monoclonal antibody produced by
CC Streptococcus mutans, particularly mouse-hybridoma MHP126 (FERM P-17566)
CC or mouse-hybridoma MHP136 (FERM P-17567), against an enzyme having
CC inhibitive activity against water insoluble glucan synthetase of glucosyl
CC transferase-B. The monoclonal antibody specifically inhibits water
CC insoluble glucan synthetase of Streptococcus mutans produced glucosyl
CC transferase-B and is used in the immunotherapy of dental caries. This
CC sequence represents a Streptococcus mutans monoclonal antibody-related
CC protein
XX
SQ Sequence 1476 AA;
Query Match 56.0%; Score 4672.5; DB 5; Length 1476;
Best Local Similarity 57.1%; Pred. No. 1.7e-274;
Matches 907; Conservative 210; Mismatches 332; Indels 139; Gaps 16;
QY 1 MEKNVRFKMHKVRWVTLVSASATMLASALGASVASADTDTA-----SDDSNOQAVVTGQ 56
DB 1 MDKVRVYKLRKVRWVTVSVASAVMTLTLSGLVKADSNESKSI SINDSNTSVVTANE 60
QY 57 TTN--NQATDQTSIAATATSEQASASTDAATDQASAAEQGTGTTASTDTAAQTNTNANEAK 114
DB 61 ESNVITEATSKQEAASSQTNHTVTSSSTSVVNPKE-----VVSNPYTVTGETASNGEKL- 115
QY 115 WYPTENENOGFTDEMLAEA-----KNVATAESDSIPSD-----LAKMSNVKQVDGKY 161
DB 116 -----QNTTVDKTSAAANNISKQTTEDTDVDDSNAAANLQILEKLPNVKEIDGY 169
QY 162 YYDDQGNVKNFVSVGDKIYIFDETGYKDTSKVDADKSSASVSONATIFAANNRAYS 221
DB 170 YYDDNNGKVRTNFTLIADGKILHFDGTAYTDSIDTVNK--DIVTTRSNLYKKYNOVD 227
QY 222 TSAKQFEAVDNYLTADSWTRPKSILKDGKTWTESGKDDRPRLMAWPDPTETKRNVTNM 281
DB 228 RSAQSFHVDHYLTAEWSRPRPKYILKDGKTWTQSTEKDFRPLMTWWPDQETQRQVNTM 287
QY 282 NKVVVGIDKTYTAETSQADLTAAELVQARIKQITSENNTKWLREAIASFVKTPQWNGE 341
DB 288 NALGLINKYTDTSNQLQUNIAAATQAKIEAKITFLKNDWLQRQTSIAFVKTSANNSD 347
QY 342 SEKPYDDHLQNGALLFDNQTLTPTDQTSNRYLLNRTPTNQTSLSRSFTYFNPNDPLGGYD 401

DB 348 SEKPFDDHLQNGAVLYDNECKLTPYANSNRYLLNRTPTNQTKKDPYV--ADNTIGGYE 405
QY 402 FLLANDVDNSNPVVOAEQLNHLNFGSIYANDADANFDSIRVDVNDVADLLOISS 461
DB 406 FLLANDVDNSNPVVOAEQLNHLNFGSIYANDADANFDSIRVDVNDVADLLOIAG 465
QY 462 DYLKAAAYGIDKNNKANNHVSIVEAWSDDNTPYLHDDGDMNNDKFRLSMLWSLAKPL 521
DB 466 DYLKAAAGIHKNDKAANDHLSILEAWSDDNTPYLHDDGDMNNDKFRLSLFSLAKPL 525
QY 522 DKSGNLPLHNSLDREVDRVETVPSYSFARAHDSVQDIIRDIKAEINPNPSFGYS 581
DB 526 NQSGMNPITNSLVNRDNDNAETAAVPSYSFIRAHDSVQDLIRDIKAEINPNVVGYS 585
QY 582 FTQEEIEQAPKIYNEDELKKTCKKYTHYVNPPLSYLLLTNKGSI PRVYVYGMFTDDGGYMA 641
DB 586 FTMBEIKKAPFIYNKOLLATEKKYTHYNTALSVALLLTNKSSVPRVYVYGMFTDDGGYMA 645
QY 642 NKTWNYDAIESLLKARMKYVSGQAMQNTYOIGNEIILTSVRYGKGLKQSKDGDATTRTS 701
DB 646 HKTINYEAEITLLKARIKYVSGQAMRNQGVNSEIITSVRYGKGLKATDGTGRTTTS 705
QY 702 GVGVMGNQPNFSLDGK--VVALNMGAAHANQERYALMVSITKGVATYATDADASKAGLVK 760
DB 706 GVAVIEGNNPSSLRLKASDRVVMNGAAHKNOAYRPLLLTTDNGIKAYHSDQEA--AGLVR 763
QY 761 RTDENGYLFLNDDLKGVANPQVSGFLOVWPVGAADDQDQIRVAASDASTDGSLSHODA 820
DB 764 YTNDRGELIFTAADIKGYANPQVSGYLGVWPVGAADQDQIRVAASDASTDGSLSHODA 823
QY 821 AMDSRVWFSGFNFQSFATKEBEYTNVIANNVDFKFSWGITDFEMAPQVVSSTDDGFLD 880
DB 824 ALDSRVWFSGFNFQSFATKEBEYTNVIANNVDFKFSWGITDFEMAPQVVSSTDDGFLD 883
QY 881 SVIQNGYAFTRDYDLGMSKANKYGTADQLVKAIKALHAKGLKVMADWVPDQMTFFPKQEV 940
DB 884 SVIQNGYAFTRDYDLGMSKANKYGTADQLVKAIKALHAKGLKVMADWVPDQMTFFPKQEV 943
QY 941 VTVTRTDKPKIAGSQINHSILYVTTKSSGDDYQAKYGGAFLDELKEKYPELPKTKQIS 1000
DB 944 VTATRVDTKYGTFTVAGSQIKNTLYVVDGKSSGKQQAQYGGAFLEELQAKYPELPKTKQIS 1003
QY 1001 TQOAPDPSVKIKOWSAKYFNGSNILGRGADYVLSDOVSNNKYFNVA--SDTLFELPSSILGK 1058
DB 1004 TGVPWDPSPVKIKOWSAKYFNGSNILGRGADYVLSDOVSNNKYFNVA--SDTLFELPSSILGK 1063
QY 1059 VYESGIRYDGKGIYINSSATGQVKAFTTEAGNLYYFGKDGVMVTGAQTINGANYPFLE 1118
DB 1064 DSQVGFSDYDGKGVYVYST--SGYQAKNTFISEGDKWYTFDNGYVMTGAQSGINGVNYFSL 1122
QY 1119 NGTALNTIYTDQGNSHYANDGKRYNGYQOF--GNDWRYFKDGMNVAGLTITVDGNVQY 1177
DB 1123 NGLQLRDAILKNEDEGTIAYYNGDGRYENGYYQFMSGVWRHFNNGEMSVGLTVIDGVQY 1182
QY 1178 FPKDGVQAKDLIVTRDQKRVYFDQHGNAATFTIADKTGHYLYLKGQVAVTGAQTVG 1237
DB 1183 FPMGVOAKGKFTVTTADGKIRYFDKSGNWRNRFTEBEGRWLYLGEDGAAVTGQITIN 1242
QY 1238 KQKLYFEANGQVKGDFVTSDEGLYFYDVDSGDMWTDTFIEDKAGNWFYLGKDGAAVTG 1297
DB 1243 GQHLFYFRANGVQVKGFEVTDHGRISYYDNGSDQIRNRFVRNAQOQWYFDPNNGYAVTG 1302
QY 1298 AOTIRGOKLYFKANGQVKGDI VKGTDGKIRYDAKSGQVFNKTVKAADGKTYVTGNDG 1357
DB 1303 ARTINGOHLFYFRANGVQVKGFEVTDHGRISYYDNGSDQIRNRFVRNAQOQWYFDPNNG 1362
QY 1358 VAVDPSVVKQTFKDSAGALRYNLKQLVTGSGWVETANHDWVYIQSKALTEGOTING 1417
DB 1363 YAV-----TGARTING 1373
QY 1418 QHLYFKEDGHQVKGQVLTGTDGKRVYDANSGDQAFNKSVTYVNGKTYFENDGCTAOTAGN 1477

Db 1374 QHLYFRANGVQVKGFEVTDHGRISYYDNGSGDQIRNRFV----- 1413
 QY 1478 PKGQTFKGDSDIRFYSMEGQLVTGSGWYENAAQOQWLYV- KNGKVLRLGLTVGSRVYFDE 1536
 Db 1414 -----RNAQOWFYFNNGVAVTGARTINGQHLFYRA 1445
 QY 1537 NGIOAKGKAVRTSDGKIRYFDENSGSMI 1564
 Db 1446 NGVQVKGFEVTDYGRISYYDANSGRV 1473

RESULT 5

AAU98027
 ID AAU98027 standard; protein; 1475 AA.

AC AAU98027;
 XX 1

DT 27-AUG-2002 (first entry)

XX S. mutans glucosyltransferase GTFB.

DE Glucosyltransferase; GTFB; transgenic plant; paper sizing;

KW coating composition; glucan; starch; latex; thermoplastic molecule;

KW amyloplast; vacuole; paper manufacture.

XX Streptococcus mutans.

OS US2002031826-A1.

PN 14-MAR-2002.

XX 19-DEC-2000; 2000US-00740274.

XX 07-JUN-1995; 95US-00478704.

PR 07-JUN-1995; 95US-00482711.

PR 07-JUN-1995; 95US-00485243.

PR 16-JAN-1998; 98US-00007999.

PR 16-JAN-1998; 98US-00008172.

PR 20-JAN-1998; 98US-00009620.

PR 11-DEC-1998; 98US-00210361.

XX (NICH/) NICHOLS S E.

XX Nichols SE;

XX WPI; 2002-414332/44.

DR N-PSDB; ABK52938.

XX Glucosyltransferase B or D protein useful for producing a glucan useful

PT as substitutes for and additions to modified starch and latexes in paper

PT manufacture, comprises mutations in specific positions.

XX Disclosure; Page 21-25; 44pp; English.

CC The invention an isolated protein comprising a glucosyltransferase (GTF)

CC B polypeptide having changes at position from I48V, D457N, D567T,

CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T,

CC I448Y/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a

CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,

CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the

CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its

CC complementary polynucleotide, a ribonucleic acid sequence encoding the

CC GTF mutant, an expression cassette comprising the polynucleotide operably

CC for producing a glucan in a plant. The method comprises transforming a
 CC plant cell with the vector, growing the plant cell under plant growing
 CC conditions to produce a regenerated plant and inducing expression of the
 CC polynucleotide for a time sufficient to produce the glucan in the
 CC regenerated plant, where the vector contains a transit sequence from
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.
 CC Glucans are useful as substitutes for and additions to modified starch
 CC and latexes in paper manufacture. Unlike prior art techniques, which
 CC require input materials that produce chemical effluents, paper
 CC manufacture utilising the glucan produced by GTF, which utilises
 CC biologically produced input materials, is more cost-effective and
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
 CC properties and impart gloss to the paper during coating step. The present
 CC sequence represents GTFB
 XX
 SQ Sequence 1475 AA;

Query Match 54.9%; Score 4580; DB 5; Length 1475;

Best Local Similarity 56.4%; Pred. No. 7e-269;

Matches 896; Conservative 212; Mismatches 338; Indels 144; Gaps 18;

QY 1 MEKNVRFKMHKVKRWVTLVSASATMLASALCASVASADTDTA----SDDSNOAVVTGQ 56

Db 1 MDKVRVYKLRKVKRWVTVSVASAVMTLTLSGGLVKADSNESKQISNDSNTSVVTANE 60

QY 57 TTN--NOATDOTSIAATATSEQASTDAATQASAAEQTGTASTDTAAQTNNANEAK 114

Db 61 ESNVITEATSKQEAASSQINHTVTTSSSTSVVNPKE----VVSNPYTVGETASNGEKL- 115

QY 115 WYPTENQGFDEMLAEA-----KNVATAESDIPSD-----LAKSNVVKQVDGKY 161

Db 116 -----QNGTTVDKTSEAAANNISKQTTAEDTVIDDSNAANLQLEKLPNKVEDGKY 169

QY 162 YYYDODGNVKNFAVSGDKIYVFDETGAYKDTSKVDADKSSVASVSONATIFAANNRAYS 221

Db 170 YYDNNGKVRTNFTLIADGKILHFDETGAYDTSDTVNK--DIVTTRNLYKYNQVVD 227

QY 222 TSAKNFEAVDNYLTADSWYRPKSILKDGKTWTESGKODFRPLLMAWPPDTETKRNYYNM 281

Db 228 RSAQSFEHVDHYLTAESWYRPKYILKDGKTWTQSTEKDFRLLMTWPPDQETQRQVNYM 287

QY 282 NKVVGIDKTYTAETSQADLTAAAEIVQARIKQITSENNTKWLREISAFAVKTPQWNGE 341

Db 288 NAQLGINKTYDDTSNQLQNLIAAATIAQAKIATTLKNTDWLRQTISAFVKTQSAWNSD 347

QY 342 SEKPYDDHLQNGALLFDNOTDLTPTQSNVRLNRTPTNQTGSLDSRFTYNNPDPLGGYD 401

Db 348 SEKPFDDHLQNGAVLYDNEGKLTTPYANSNYRLNRTPTNQTGKDPRT--ADNTIGGYE 405

QY 402 FLLANDVDNSNPFVQAEQLNWLHYLLNFGSIYANDADANFDSIRVADVNDVADLLQISS 461

Db 406 FLLANDVDNSNPFVQAEQLNWLHFLMNFNFIYANDPDANFDSIRVADVNDVADLLQIAG 465

QY 462 DYLLKAAVIGDKNKAANNHVSIVEASDNDTPLYLHDDGDLNMMNKKPLSLMLWSLAKPL 521

Db 466 DYLLKAAVIGDKNKAANNHVSIVEASDNDTPLYLHDDGDLNMMNKKPLSLMLWSLAKPL 525

QY 522 DKRSGLNPLIHLNSLVDREVDREVEVTPSYSPARAHDSVQDIIRDIKAEINPNPSFGYS 581

Db 526 NQSGNPNLITNSLVNRTDDNDAETAAVPSYFIRAHDSVQDDIADIKAEINPNVVGYS 585

QY 582 FTQEEIEQAFKINEDLKKTKKYTHYNNVPLSYTLTLLTNKGSIPRVYVYGMDFDDGQYMA 641

Db 586 FTMEIEKKAFEIYNKOLLATEKKYTHYNTALSVALLLTNKSSVPRVYVYGMDFDDGQYMA 645

QY 642 NKTVNYDAIESLLKARKVYSGQAMQNYQINGEILLTSVRYGKGLKSKDKGDAVTRTS 701

Db 646 HKTINYEALETLLKARIKYVSGQAMRNQOVGNSEIITSVRYGKGLKATDTGDRTRTS 705

QY 702 GGVVYMGPNPFLSDGK-VVALNMGAAHAAHQEYRALMVSTKQGVATYATDADASKAGLVK 760

Db 706 GVAIEGNNPSELKASDRVVVMGAHKNQAYRPLLLTTDNGIKAYHSQDEA--AGLVR 763
QY 761 RTDENGYYFLNDDLKGVANPOVSGFLQVWPVGAADDQDIRVAASDTASTDGK--SLHQ 818
Db 764 YTNDRGELIFTAADIKGYANPOVSGYLGWVWPVGA---LTKFALRLARPHQOMASVHQ 820
QY 819 DAAMDSEVRPEGSNFSQSFATKEEYTNVVIANNVDFKFSVSGITDPEMAPQYVSSTDGQF 878
Db 821 NAALDSRVMEGFSNFQAFATKEEYTNVVIANNVDFKFAEWGVTDPEMAPQYVSSTDGSF 880
QY 879 LDSVIQNGYAFTDRYDLGMSKANKYGTADOLVKAIRKALHAKGLKVMADWVPDQWYTFPKQ 938
Db 881 LDSVIQNGYAFTDRYDLGMSKANKYGTADOLVKAIRKALHAKGLKVMADWVPDQWYTFPKQ 940
QY 939 EVTVTRTRDKFGPIAGSQINHSLYVTDTKSSGDDYQAKYGGAFDLDELKYEPELFTKKQ 998
Db 941 EVTVTRTRDKFGPIAGSQINHSLYVTDTKSSGDDYQAKYGGAFDLDELKYEPELFTKKQ 1000
QY 999 ISTQOAIIDPSVKIKQSAKYFNGSNILGRGADYVLSQVSNKYFNVA--SDTLFLPSSLL 1056
Db 1001 ISTQOAIIDPSVKIKQSAKYFNGSNILGRGADYVLSQVSNKYFNVA--SDTLFLPSSLL 1060
QY 1057 GKVVESGIRYDGRGKYLINSSATGDQVKASPIEAGNLYYFGKDGVMVTGAOTINGANYFP 1116
Db 1061 NQDSQVGFSDGKGYVYIST--SGYQAKNTPISEGRKWIYFDNNGYVMTGAQSGINGVNYF 1119
QY 1117 LENGTLRNTIYDAQNSHYANDGKRYENGQQF--GNDWRFKDGNMVAVLGLTVDGNV 1175
Db 1120 LSNGLQLRALDKNEVDGTAYYNGDGRYENGYYQFMSGWRHFNGEMSVGLTVIDGQV 1179
QY 1176 QYFDKQGVQAKKIIVTRDGKVRYPHOHNGNAATNTFIADKTHWHYLGKGVAVTGAQT 1235
Db 1180 QYFDKQGVQAKKIIVTRDGKVRYPHOHNGNAATNTFIADKTHWHYLGKGVAVTGAQT 1239
QY 1236 VQKQKLYFRANGQVKGDFVTSDEGLYFYVDVDSGDWMTDTTFIEDKAGNWFYLGKGAAY 1295
Db 1240 INQHLYFRANGQVKGDFVTSDEGLYFYVDVDSGDWMTDTTFIEDKAGNWFYLGKGAAY 1299
QY 1296 TGAQTIRGOKLYFKANGQVKGDIKGTGKIRYDAKSGEQVFNKVAADGKTIVIGN 1355
Db 1300 TGARTINGQLLYFRANGQVKGDFVTSDEGLYFYVDVDSGDWMTDTTFIEDKAGNWFYLGKGAAY 1359
QY 1356 DGVAVDPSVVGKGTFKDASGALRYFNKLGQLVGTSGWYETANHDWYIYOSKALTCGEQTI 1415
Db 1360 NGYAV-----TGARTI 1370
QY 1416 NGQHLFYKEDGHQVKGQLVGTGDKVRYVDANSQDQAFNKSVTYNGKTYFNGDGTQATA 1475
Db 1371 NGQHLFYFRANGQVKGDFVTSDEGLYFYVDVDSGDWMTDTTFIEDKAGNWFYLGKGAAY 1412
QY 1476 GNPKGQTFKDGSDIRFYSMEGQLVTGSGWYENAGQWLVY--KNGKVLTLGLQTVGSRVYF 1534
Db 1413 -----RNAQGGWFFDNGYAVTGARTINGQLLYF 1442
QY 1535 DENGIOAKGAVRTSDGKIRYFDENSGSMI 1564
Db 1443 RANGVQVKGDFVTSDEGLYFYVDVDSGDWMTDTTFIEDKAGNWFYLGKGAAY 1472

RESULT 6

ADD93654
ID ADD93654 standard; protein; 1475 AA.
XX
AC ADD93654;
XX
DT 29-JAN-2004 (first entry)
XX
DE Streptococcus mutans glucosyltransferase-B.
XX
KW Glucosyltransferase; enzyme; vaccine; anticonvulsant; epitope; immunogen.
XX
OS Streptococcus mutans.
XX

PN WO2003075845-A2.
XX
XX 18-SEP-2003.
XX
XX 07-MAR-2003; 2003WO-US006962.
XX
XX 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
XX
XX (FORS-) FORSYTH INST.
XX
XX Smith DJ, Taubman MA;
PI WPI; 2003-845091/78.
XX
XX Composition useful as vaccines for dental caries comprises a fragment of
PT a glucan binding protein-B binding to a major histocompatibility complex
PT class II protein.
XX
XX Claim 16; Page 12-13; 49pp; English.
XX
CC The present sequence is the protein sequence of Streptococcus mutans
CC glucosyltransferase-B (GTF-B). Peptide fragments of GTF-B, especially
CC from the catalytic domain of the polypeptide, can be used in immunogenic
CC compositions and subunit vaccines for dental caries. These compositions
CC comprise a major histocompatibility complex (MHC) class II protein-
CC binding peptide from S. mutans glucan binding protein-B (GbpB)
CC covalently linked with a peptide fragment of a streptococcal
CC glucosyltransferase. The compositions are used in a claimed method of
CC eliciting production of an antibody in a mammal. Diepitopic or
CC multi-epitopic polypeptides can be prepared synthetically or by
CC recombinant DNA technology. Antibodies raised against MHC class II
CC binding fragments of GbpB can be used in passive immunisation.
XX
SQ Sequence 1475 AA;
Query Match 54.9%; Score 4580; DB 7; Length 1475;
Best Local Similarity 56.4%; Pred. No. 7e-269;
Matches 896; Conservative 212; Mismatches 338; Indels 144; Gaps 18;
QY 1 MEKVRFPKHKYKRWVTLVSASATWLASALGASVASADTDFA----SDSNQAVVTGDQ 56
DB 1 MDKVRYPKHKYKRWVTLVSASATWLASALGASVASADTDFA----SDSNQAVVTGDQ 56
QY 57 TTN--NOATDQTSIAATATSEQASATDAATDQASAAEQGTGTTASTDTAAQTITNANEAK 114
DB 61 ESNVTEATSKQEAASSQTNHTVTTSSSTSVVNPKE-----VVSNPYTVGETASNGEKL- 115
QY 115 WYPTENENQGFDEMLAE-----KNVATAESDSIPSD-----LAKMSNVKQVDGKY 161
DB 116 -----NQTTTVDKTSAAANNISKQTTEDATDVDDSNAAALQLLEKLPNVKEIDGKY 169
QY 162 YYYDQGNVKNFVAVSGDKIYYFDETGAYKDTSKVDADKSSASVSONATIFAAANNRAYS 221
DB 170 YYYDQGNVKNFVAVSGDKIYYFDETGAYKDTSKVDADKSSASVSONATIFAAANNRAYS 227
QY 222 TSANKFEAVDNYLTADSWVRPKSILKDGKTWTEGKDDPRLLMAMWPDPTETKKNVNYM 281
DB 228 RQAQFEHVDHYLTAEWSYRPRYILKDGKTWTEGKDDPRLLMAMWPDPTETKKNVNYM 287
QY 282 NKVVGIDKTYAETSQADITAAAEVLQARIEQKITSENNKWLREAIASAFVKTQPOWNGE 341
DB 288 NAQLGINKTYDDTSNQLQNLNIAAATIQAKIEKITTLKNTDMLRQTISAFVKTQSAWNSD 347
QY 342 SEKPYDDHLQNGALLFDNQDTLTPDTQSNRYLLNTPNTQTSLSRFRYPNPNPDLGGYD 401
DB 348 SEKPYDDHLQNGALLFDNQDTLTPDTQSNRYLLNTPNTQTSLSRFRYPNPNPDLGGYD 405
QY 402 FLANDVNSNPVQAEQLNMLHLYLLNFGSIYANDADANFDSIRVDADVNDADLLQISS 461
DB 406 FLANDVNSNPVQAEQLNMLHLYLLNFGSIYANDADANFDSIRVDADVNDADLLQISS 465
QY 462 DYLKAAAGIDKKNKNNHNSIVEAWSNDTPYLHDDGDNLMNDKPNRSLMSLAKPL 521

CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
CC complementary polynucleotide, a ribonucleic acid sequence encoding the
CC GTF mutant, an expression cassette comprising the polynucleotide operably
CC linked to a promoter, a vector comprising the expression cassette, host
CC cell introduced with the vector, a transgenic plant comprising the
CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
CC coating composition comprising a glucan produced in a plant transformed
CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
CC thermoplastic molecule or their combinations or glucan and starch where
CC the glucan is produced in the amyloplast and/or vacuole or a maize line
CC deficient in starch biosynthesis, transformed with a gene encoding a
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
CC comprising the glucan (paper sizing/coating agent). The vector is useful
CC for producing a glucan in a plant. The method comprises transforming a
CC plant cell with the vector, growing the plant cell under plant growing
CC conditions to produce a regenerated plant and inducing expression of the
CC polynucleotide for a time sufficient to produce the glucan in the
CC regenerated plant, where the vector contains a transit sequence from
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
CC is produced in the amyloplast of potato or the vacuole of sugar beet.
CC Glucans are useful as substitutes for and additions to modified starch
CC and latexes in paper manufacture. Unlike prior art techniques, which
CC require input materials that produce chemical effluents, paper
CC manufacture utilising the glucan produced by GTF, which utilises
CC biologically produced input materials, is more cost-effective and
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
CC properties and impart gloss to the paper during coating step. The present
CC sequence represents a GTFB mutant of the invention. Note: The present
CC sequence is not shown in the specification but was created by the indexer
CC using the GTFB sequence appearing as AAU98027 and the information in
CC claim 36
CC
XX

SQ Sequence 1475 AA;

Query Match 54.8%; Score 4579; DB 5; Length 1475;
Best Local Similarity 56.3%; Pred. No. 8.1e-269;
Matches 895; Conservative 213; Mismatches 338; Indels 144; Gaps 18;

QY 1 MEKNVRFKHKVKKRWVTLVSASATMLASALGASVASADTDTA----SDSNQAVVTGQ 56
DB 1 MDKKVYKURKKRWVTVSASAVMTLTLSSGLVKAUSNESKQISNDSNTSVVTANE 60
QY 57 TTN--NQATDQTSIAATATSEQASSTDAATQASAAEQTGGTASTDTAAQTNTANEAK 114
DB 61 ESNVITEATSKQEAASQTNHTVTTSSTSTSVVNPK-----VVSNPVTVGETASNGKL- 115
QY 115 WVPTENEGQTDMLAEA-----KNVATAESDSIPSD-----LAKSNVQVQDGKY 161
DB 116 -----QNTTVDKTSAAANNISQTTTEADTDVDDSNAAQLQLEKLPNYKEIDGKY 169
QY 162 YYYDQGNVKNFAVSGDKIYYFDETGAYKDTSKVDADKSSAVSQNATIFAAPNRRAYS 221
DB 170 YYYDNNKVRTNTLTADGKLHFDGTAYDTISIDTVNK--DIVTTRSNLYKKYNQVVD 227
QY 222 TSAKNFEAVDNYLTADSWYRPSILKDGKTWTESGKDDFRPLLMAMPDPTETKRNVTNYM 281
DB 228 RSAQSFEHVDHYLTASWYRKPVIKDGKTWTQSTTEKDFRPLMTWMPDQETQRQVYNYM 287
QY 282 NKVVGIDKTYTATSQADLTAAAEVQARIEQKITSNNYTKMLREASAFVKTOPOWNGE 341
DB 288 NAQLGINKTYDTSNQLNLNAAATIQAKEAKITLLKNTDMLRQRTISAFVKTSQAWNSD 347
QY 342 SEKPYDDHLQNGALLFDNQDLPDQTSNRYLLNRPTNQTGSLDSRFTYVNPNDPLGGYD 401
DB 348 SEKPFDDHLQNGAVLDYNEKGLPIYANSVRIINRPTNQTGKDDPRYT--ADNTIGGYE 405
QY 402 FLLANDVNSNPVQAEQLNWLHLYLNFGSIYANDADANPDSIRVDAVDNVDADLLQISS 461
DB 406 FLLANDVNSNPVQAEQLNWLHFLMFGNIYANDPDANFDSVRVDAVDNVDADLLQIAG 465
QY 462 DYLAAYGIDKXNKNHNSVIVEASMSDNTPTYLHDDGNLMMNDKFRLSMLWSLAKPL 521

DB 466 DYLAAKAGHKHDKKANDHLSILEAWSNDTPLYLHDDGNMNMNDKLSLFLSLAKPL 525
QY 522 DKRSGNLPLIHNSLDVREDDREVEFVPSYSFARAHDSVQDIIRDIIKAEINPNSFGYS 581
DB 526 NQSRGWNPLITNSLVNRDNDNAETAAPSYSFIRAHDSVQDLADIIRKAEINPNVVGYS 585
QY 582 FTQEBIEQAFKLYNEDLKKTKKYTHYNVPLSYTLILLTNKSGIPRVYVYGDMDTDDQYNA 641
DB 586 FTMBEIKKAFETYNKDLATEKKYTHYNALSYALLTNKSSVPRVYVYGDMDTDDQYNA 645
QY 642 NKTVNYDAIESLKKARMKVSGQAMQNYQIGNGEILTSVRYGKALKGSKDGKDATTRTS 701
DB 646 HKTIIYEALTEILLKARIKYVSGQAMRQOVGNSEILTSVRYGKALKALATDGTDRTRTS 705
QY 702 GGVVNGQPNFSLDGK--VVALNMGAAHANQBYRALMSTKDGVIATYATDADASKAGLVK 760
DB 706 GVAVIEGNNPSLURLKASDRVVVNMGAHKHQAAYRPLLLTTDNGIKAYHSDQEA--AGLVR 763
QY 761 RTDENGILYFLNDDLKGVANPOVSGFLQVWVPVGAADDODIRVAASDTASTDGK--SLHQ 818
DB 764 YTNDRGELIFTAADIKGVANPOVSGYLVGVWVPVGA--LIKMFALRLARPHQOQMASVHQ 820
QY 819 DAAMDSRVMEFGFSNFQSPATKEEYTNVVIANNVDKFSVSGWITDPEMAPOQVVSSTDGOF 878
DB 821 NAALDSRVMEFGFSNFQSPATKEEYTNVVIANNVDKFAEWGTVDEMAPOQVVSSTDGSGF 880
QY 879 LDSVIQNGYAFTRYDRLGMSKANKYGTADQLKAIKALHAKGLKVMADVPDQMTFFPKQ 938
DB 881 LDSVIQNGYAFTRYDRLGMSKANKYGTADDLKAIKALHAKGLKVMADVPDQMTFAPEK 940
QY 939 EYVTVTRTDKPKPTAGSQINHSILYVDTFKSSGDDYQAKYGGAFDLDELKPKPELTKKQ 998
DB 941 EYVTVTRVDKYGTPVAGSQIKNTLYVVDGSKSGKQQAQYGAFLQAKYPELFAKQ 1000
QY 999 ISTGQAIDPSVKIKOWSAKYFNGSNILGRGADVLSQDVSNKYFNVA--SDTLFLPSSLL 1056
DB 1001 ISTGVPMDSVKIKOWSAKYFNGTNILGRGAGYVLKQDQATNTYFNISDNKEINFLPKTLL 1060
QY 1057 GKVBESGTRYDGKGYIYNSSATGDQVKASFIPEAGNLHYFGKDGVMVTGAQTINGANYFF 1116
DB 1061 NQDSQVGFSDGKGYVYVYST-SGYQAKNTFISEGDKWYVFDNNGYVMTVQAQSGINGVNYF 1119
QY 1117 LENGTAIRNTIYTDAGNSHYANDGKRYENGYQOP-GNDWRYFKDGNMAGVLTGTVDGNV 1175
DB 1120 LSNGLQLRDAILKNEIDGTAYYVNDGRRYENGYQPMGVRHFNHNGEMSVGLTVIDGQV 1179
QY 1176 QYFDKDGQVAKDKIIIVTRDGKVRYFDQHNAGNAATNTFIADKTHWYLYLKGDKGVAVTGAOT 1235
DB 1180 QYFDEMGYQAKGFVTTADGKIRYFDKQSGNMRYNRFIENEBSKWLXLYLGEDGAATVGSOT 1239
QY 1236 VGKQKLYFEANGQVKGDFVTSDEGKLYFYVDVSDGDMWTDFTFIEDKAGNWFYLGKGAUV 1295
DB 1240 INGQHLVFRANGVQVKGFEVTDHHRISYVDGNSGQDIRNRFVRNAQOGVYFDNNGYAV 1299
QY 1296 TGAQTRIGQKLYFKANGQVKGDIYKGTGDKIRYDYDAKSGEQVFNKTVRAADKTYVIGN 1355
DB 1300 TGAQTINGQLLYFRANGVQVKGFEVTDYGRISYVDGNSGQDIRNRFVRNAQOGVYFDN 1359
QY 1356 DGVAVDPSVVGQTFPKDASGALRFYNLKGQLVTGSGWYETANHDWVVIQSGKALTGBQTI 1415
DB 1360 NGYAV-----TGARTI 1370
QY 1416 NGQHLVFKEDGHQVKGQLVTGDKVRYDYDANSQDQAFNKSVTNKGTYVYFGNDGTAQTA 1475
DB 1371 NGQHLVFRANGVQVKGFEVTDHHRISYVDGNSGQDIRNRFV----- 1412
QY 1476 GNPQGTQFGGSDIRFYSMEGQLVTGSGWYENACQWLYV-KNGKVLTLQTVGSQRYVF 1534
DB 1413 -----RNAQOGVYFDNNGYAVTGARTINGOHLXYF 1442
QY 1535 DENGIOAKGKAVRTSDGKIRYFDENSGSMI 1564

Query Match 54.8%; Score 4575; DB 5; Length 1475;
Best Local Similarity 56.3%; Pred. No. 1.4e-268;
Matches 895; Conservative 213; Mismatches 338; Indels 144; Gaps 18;

QY 1 MEKNRPFKHKVKKRWTTLSVASATMLASALGASVASADTDTA-----SDDSQAQVVTGDO 56
DB 1 MDKVRYLKRVKRWTVSVASAVMTLLTSLGGLVKADSNESKQSISNDSNTSVVTANE 60

QY 57 TTN--NOATDQTSIAATPATSEQASDAADQASAAEQTGCTTASTDTAAQTATTNANEAK 114
DB 61 ESNVITEATSKQBAASSQTNHTVTSSSTSVVNPKE---VVSNPYTVGETASNGEKL- 115

QY 115 WVTFENENQGTDEMLAEA-----KVVATAESDSIPSD-----LAKMSNVKQVDGKY 161
DB 116 -----QNQTTTVDKTSAAANNISKQTTTEADTDVDDSNAAQLILEKLPNKEIDGKY 169

QY 162 YYDDQGNVKNFAVSGDKIYFDETGAYKDTSKVDADKSSASVSNATIPANNPAYS 221
DB 170 YYDNNKVRNTFTLIADGKILHFDEGTGATDTSIDTVNK--DIVTTRSLYKXNQVYD 227

QY 222 TSAKNFEADVNYLTADSWYRPKSILKDGKTWTBESGDDFRPLLMAMPDPTETKRNYYNM 281
DB 228 RSAQSFEHVDHLYLTAESWYRPKYLKDGKTWTOSTEKDFRPLMTWHPDOETQRYNYM 287

QY 282 NKVVGIDKTYTASTSQADLTAAELQVAREIQKITSNNNTKWLREAIASFVKTPQWNCE 341
DB 288 NAQGLINKTYDTSNQLNLIAAATIAQIEAKITLLKNTDLWLRQTISAFVKTSANNSD 347

QY 342 SEKPYDHLQNGALLFQNTDLPTQSNVRLNRTPTNQTGSLDSRFTYNNPNDPLGGYD 401
DB 348 SEKPFDDHLQNGAVLYDNEGLTPYANSNRYILNRTPTNQTGKDKPRYT--ADNTIGGYE 405

QY 402 FLIANDVDSNPPVQAEQLNHLNLFNFGSIYANDADANFDSIRVDADVNDVADLLQISS 461
DB 406 FLIANDVDSNPPVQAEQLNHLNLFNFGSIYANDADANFDSIRVDADVNDVADLLQIAG 465

QY 462 DYKAAAGIDKNNKNNANNHVSIVEAWSNDPTPYLHDDGNLMMNDKFLRSLMWSLAKPL 521
DB 466 DYKAAAGIHKNDKAANDHLSILEAWSNDPTPYLHDDGNMINDNKLRLSLFLSLAKPL 525

QY 522 DKRSGLNPLIHLNSLVREVDREVEVPSVSPARAHDSSEVODIIRIIKAEINPNPFGVS 581
DB 526 NORSNGNPLITSLNVRNTDQNAETAAPVPSYFIRAHDSSEVODIIRIIKAEINPNVVGVS 585

QY 582 FTQEEIEQAFKIYNEDLKKTKKYTHYNVPLSLYTLTLTKNGKIPRVYVYGMFTDDGQYMA 641
DB 586 FTWEEIKKAFIYNKDLLEKTKYTHYNTALSALLTKNSVPRVYVYGMFTDDGQYMA 645

QY 642 NKTVNYDAIESLLKARMKYVSGQAMQNYQIENGELITSVRYGKALKOSDKGDATRTS 701
DB 646 HKTINYEAETLLKARIKYVSGQAMRNQVGNSEIITSVRYGKALKATDTGDRITRTS 705

QY 702 GYGVVMGQNPNSLDGK--VVALNMGAAHNOEYRALMVSTKQGVATYATDADASKAGLVK 760
DB 706 GVAIVTEGNNPSRLKASDRVVMNGAAHKNQAYRPLLLTTDNGIKAYHSDQEA--AGLVR 763

QY 761 RTDENGVLFLNDLKGAVNPQVSGFLQVWPVGAADDQDQIRVAASDTASTDGK--SLHQ 818
DB 764 YTNDRGELLFTAADIKGYANPQVSGYLVWPVGA---LTKMFAURLARPHQOMASVHQ 820

QY 819 DAAMDSRVMEFGFSNFQSPATKEEYTNVVIANNVDKFSWIGITDPEMAPQVYSSTDGQF 878
DB 821 NAALDSRVMEFGFSNFQATKKEEYTNVVIANNVDKFAEWGVTDFEMAPQVYSSTDGSP 880

QY 879 LBSVQNGVAFTRDYRDLGMSKANKYGTADQLKAIKALHAKGLKVNADWVPQMYTFPKO 938
DB 881 LBSVQNGVAFTRDYRDLGMSKANKYGTADQLKAIKALHAKGLKVNADWVPQMYTFPEK 940

QY 939 EVVTVTRTDKFKPTAGSIOINHSLYVTDTKSSGDDYQAKYGGAFDLDELKYPELTKKO 998
DB 941 EVVTVTRTDKFKPTAGSIOINHSLYVTDTKSSGDDYQAKYGGAFDLDELKYPELTKKO 1000

QY 999 ISTGQAIIDPSVKIKQWSAKYFNGSNILGRGADYVLSQVSNKYFNVA--SDTLFLPSSLL 1056

Db 1001 ISTGVPMDPSVKIKQWSAKYFNGSNILGRGADYVLSQVSNKYFNVA--SDTLFLPSSLL 1060

QY 1057 GKVVESGIRVDGKGIYINSSATGDOVKASFTTEAGNLYYFKGQGYMTGAQTINGANYFF 1116

DB 1061 NQDSQVGSYDGGKGYVYST-SGYQAKNTFISEGDKRYFYDNNGYMTGAQSINGVNYFF 1119

QY 1117 LENGTALENTIYDAQGNSHYANDGKRYENGQQF--GNDWRYPKDCGNMAVGLTTVDGNV 1175

DB 1120 LSNGLQLRDALLKNEGDGTAYAYGNDGRRYENGYYQFMSGVWRHPNNGEMSGLTVIDGOV 1179

QY 1176 QYFDKQGVQAKKIIIVTRDGKRYRFDHNGNAATNTFIADKTHWYLYGKDGVAVTGAQT 1235

DB 1180 QYFDEMGYQAKGFVTTADGKIRYFDKQSGNMYRNRPIENEEGKWLVLGSDGAAVTGSQT 1239

QY 1236 VGKQKLYFEANGQOVKGFVTSDBGKLYFYDVSQDMWTTFTFIEDKAGNWFYLGKDAV 1295

DB 1240 INGQHLVFRANGVQVKGFEVTDHHRISYYDNGSGDQIRNRFVRNAQOGWFYFDNNGYAV 1299

QY 1296 TGAQTIRGQKLYFKANGQOVKGDIVKGTGKIRYDAKSQGEQVFNKTVKAADGKTYVIGN 1355

DB 1300 TGARTINGQLLYFRANGVQVKGFEVTDHHRISYYDNGSGDQIRNRFVRNAQOGWFYFDN 1359

QY 1356 DGVAVDPVSVVKGQTFKDGASGALRPFYNLKGQLVTGSGWYETANHDWYVIQSGKALTGBQTI 1415

DB 1360 NGYAV-----TGARTI 1370

QY 1416 NGQHLVFKEDGHQVKGQLVTGTDGKRYVYDANSQDQAPKSVTVNGKTYFYFGNDGTAQTA 1475

DB 1371 NGQHLVFRANGVQVKGFEVTDHHRISYYDNGSGDQIRNRFV----- 1412

QY 1476 GNPKGQTFKDGSDIRFYSMEGQLVTGSGWYENAGQOWLYV-KNGKVLTLQTVGSGORVYF 1534

DB 1413 -----RNAQOGWFYFDNNGYAVTGARTINGQHLYP 1442

QY 1535 DENGIOAKKAVRTSDGKIRYFDENSMSI 1564

DB 1443 RANGVQVKGFEVTDHHRISYYDANSGERV 1472

RESULT 11
AAU98033
ID AAU98033 standard; protein; 1475 AA.
XX AAU98033;
AC AAU98033;
XX 27-AUG-2002 (first entry)
DT
XX S. mutans glucosyltransferase GTFB mutant K1014T.
DE
XX Glucosyltransferase; GTFB; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture; mutant; mutein.
OS Streptococcus mutans.
OS Synthetic.
XX
XX Key Location/Qualifiers
FH Misc-difference 1014 /note= "Wild-type Lys substituted by Thr"
FT
XX US2002031826-A1.
XX 14-MAR-2002.
XX 19-DEC-2000; 2000US-00740274.
XX 07-JUN-1995; 95US-00478704.
PR 07-JUN-1995; 95US-00482711.
PR 07-JUN-1995; 95US-00485243.
PR 16-JAN-1998; 98US-00007999.
PR 16-JAN-1998; 98US-00008172.
PR 20-JAN-1998; 98US-00009620.

Qy	1296	TGAOTIRGOKLYFRANGOOVKGDIVKGTGDKIRYYDAKSGBOVENKTKVKAADGKTYVIGN	1355
Db	1300	TGARTINGQLLYFRANGVQVKGEFVTDYRGISYDNGSGDQIRNFRVRNAGQOWFVFDN	1359
Qy	1356	DCVAVIDPSVKQGTFFKQASGALRFINLKGQLVTGSGWYETANHDMWYIQSGKALTGBOTT	1415
Db	1360	NGYAV-----TGARTI	1370
Qy	1416	NGOHLYPEKEDGHQVKGLVTGTDGKVRYYDANSQDAFNKSVTVNGKTYVFGNDGTQAQTA	1475
Db	1371	NGOHLYPEFRANGVQVKGEFVTDHGRISYDNGSGDQIRNFRV-----	1412
Qy	1476	GNPKGQTFKQGSDDIRFYSMEGQQLVTGSGWTENAGQWLYV-KNGKVLTLGQTVGSQRVYF	1534
Db	1413	-----RNAGQWFFYFDNNGYAVTGARTINGQHLFY	1442
Qy	1535	DENGIOAKGKAVRTSDGKIRYFDENSGSMI	1564
Db	1443	RANGVQVKGEFVTDYRGISYDANSGERV	1472

RESULT 12

AAU98032
ID AAU98032 standard; protein: 1475 AA.

AAU98032;

DT 27-AUG-2002 (first entry)

DE S. mutans glucosyltransferase GTFb mutant D567T.

Glucosyltransferase; GTFB; transgenic plant; paper sizing;
 coating composition; glucan; starch; latex; thermoplastic molecule;
 amyloplast; vacuole; paper manufacture; mutant; mutin.

OS Streptococcus mutans.

OS Synthetic:

Key	Location/Qualifiers
-----	---------------------

FT Misc-difference 567

FT /note= "Wild-type Asp substituted by Thr"

PN US2002031826-A1.

PD 14-MAR-2002.

19-DEC-2000: 2000US-00740274.

PR 07-JUN-1995: 95US-00478704.

PR 07-JUN-1995; 95US-00482711.

PR 07-JUN-1995; 95US-00485243.

PR 16-JAN-1998; 198US-00007999.

PR 16 -JAN-1998; 98US-00008172.

PR 20-JAN-1998; 98US-00009620.
PR 11-DEC-1998; 98US-00310361

PR 11-DEC-1998; 98US-00210361.
XX

PA (NICH/) NICHOLS S E.

XX
PI
Nichols SE:

WPI: 2002-414332/44.

Glucosyltransferase B or D protein useful for producing a glucan useful as substitutes for and additions to modified starch and latexes in paper manufacture. Comprises mutations in specific positions

PS Claim 36: page: 44pp: English

CC The invention an isolated protein comprising a glucosyltransferase (GT)

CC B polypeptide having changes at position from I448V. P457N P567T.

CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T.

CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a

GTF D polypeptide having changes at positions from T599D, T589E, N471D, N471D/T589D, and N471D/T589E. Also included are a glucan produced by the GTF mutant, an isolated polynucleotide which encodes p1 or p2, or its complementary polynucleotide, a ribonucleic acid sequence encoding the GTF mutant, an expression cassette comprising the polynucleotide operably linked to a promoter, a vector comprising the expression cassette, host cell introduced with the vector, a transgenic plant comprising the vector, a seed or tuber from the transgenic plant, a paper sizing and/or coating composition comprising a glucan produced in a plant transformed with a gene encoding the mutant GTF, wild type or, starch, a latex, thermoplastic molecule or their combinations or glucan and starch where the glucan is produced in the amyloplast and/or vacuole or a maize line deficient in starch biosynthesis, transformed with a gene encoding a glucosyltransferase B or D enzyme, wild-type or mutant and a paper comprising the glucan (paper sizing/coating agent). The vector is useful for producing a glucan in a plant. The method comprises transforming a plant cell with the vector, growing the plant cell under plant growing conditions to produce a regenerated plant and inducing expression of the polynucleotide for a time sufficient to produce the glucan in the regenerated plant, where the vector contains a transit sequence from ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and chlorophyll AB binding protein to produce a transgenic plant, and glucan is produced in the amyloplast of potato or the vacuole of sugar beet. Glucans are useful as substitutes for and additions to modified starch and latexes in paper manufacture. Unlike prior art techniques, which require input materials that produce chemical effluents, paper manufacture utilising the glucan produced by GTF, which utilises biologically produced input materials, is more cost-effective and environmentally friendly. Moreover, glucans also exhibit thermoplastic properties and impart gloss to the paper during coating step. The present sequence represents a GTFB mutant of the invention. Note: The present sequence is not shown in the specification but was created by the indexer using the GTFB sequence appearing as AAU98027 and the information in claim 36

Query Match	54.8%	Score 4573;	DB 5;	Length 1475;
Best Local Similarity	56.3%	Pred. No. 1.9e-268;		
Matches 895; Conservative	212;	Mismatches 339;	Indels 144;	Gaps 18

```

Qy 1 MEKVVRFBMHVKVGRWTVLSVASATMLASALGASVAGADTDTA-----SDDSNQAVVTGDQ 56
Db 1 MDKVRVYLRVKVGRWTVLSVASAVMTLTTLGSGLVASASNESKQISNDNTSVYTTANE 60
Qy 57 TTN--NOATDQTSIAATATSEQASASTDAATDOAGAAEQTCQTASTDTAQTATTNANEAK 114
Db 61 ESNVITEATSKQEAASSQTNHTVTYSSSSTSVVPKE-----VVSNPYTVGTASNGEKL- 115
Qy 115 WVPTENENQGTDEMLABE-----KNVATAESDISPSD-----LAKMSNVKQVGDGY 161
Db 116 -----QNQTTTVDVTKTSEAAANNISKQTEADTVIDDSNAANLIQILEKLPNVKEIDGY 169
Qy 162 YYYDQDGNVKNFAVSGDKYKVFDETGAYKOTSKVDADKSSSAVSQNAATIFAANNRAYS 221
Db 170 YYDNNNGKVRNFTLIADGKILHFDETGATYDTSIDTVN:::---:::---::: 227
Qy 222 TSAKNFEADVNYLTADSWYRPKSILKDGKTWTESGKODFRPLLMAWMPDTEKRNYNVM 281
Db 228 RSAQSPHEVDHYLTAEBSWYRPFVKILDKGKTWTQSTEXDKFRPLMTWMPDQETQRYNYM 287
Qy 282 NKVVGIDKTYTAEQTSQADLTAAAEVLVQARTEOKTISENNKWLBEATISAFVKTOPOWNGE 341
Db 288 NAQLGINKTYDDTSNQLQNLNAAATIQAKTEAKITLTKNTDNLKQTTISAFVKTOSAWNSD 347
Qy 342 SEKPYDDHLQNGALLFNDQTLDPDTQSNRYLLNRTPNTQGSLDSTFTYPNPDPLGGYD 401
Db 348 SEKPFDDHLQNGAVLYDNECKLIPYANSNYRILNRTPNTQTKKDPYTT--ADNTIGGYE 405
Qy 402 FLLANVDNSNPVVQAEQLNWLHLYLNFSGSIYANDADANFDSIRVDADVNDVADLLOISS 461
Db 406 FLLANVDNSNPVVQAEQLNWLHFNFMFGNIYANDPDANFDSIRVDADVNDVADLLOIAG 465

```

QY 462 DYLKAAAGIDKNNKANNHVSIVEARSDNDTPYLHDDGDNLMNMONKPELSMLWSLAKPL 521
DB 466 DYLKAAAGIHKNDKAANDHLSILEAWSDNDTPYLHDDGDNMNMNMONKLSLFLSLAKPL 525
QY 522 DKRSGLNPLTHNSLVDREVDPVPSYSFARAHDSVQDIIRDIKAEINPNPSFGYS 581
DB 526 NQSGNPNPLTHNSLVRNDDNAETAAPVPSYSPIRAHDSVQTLIADIIRKAEINPNVVGYS 585
QY 582 FTQEEIEQAFKLYNEDLKTDKKYTHYNVPLSYTLTLLTNKGSIPRVYYGDMFTDDGQYMA 641
DB 586 FTMBEIKKAFIYNKOLLATEKKYTHYNTALSVALLLTNKSSVPRVYYGDMFTDDGQYMA 645
QY 642 NKTVNYDAIESILLKARMKVTSGGQAMONYOINGEILITSVRCKGKALQSDGDAATTRS 701
DB 646 HKTINYEALETLLKARIKVSGGQARNQOVGNSEIITSVRYGKALKATDTRGDRTRRS 705
QY 702 GGVVWGNOPNPSLQDK--VVALNMGAAHANQEVRLMVTSTKQGVATYATDADASKAGLVK 760
DB 706 GVAVIEGNPNPSLRLKASDRVNVNMGAAHKNQAYRPLLLTTDNGIKAYHSDQEA--AGLVR 763
QY 761 RTDENGILYFLNDDLKGVANPQVSGFLQVVPVGAADDQDIIRVAASDTASTDGK--SLHQ 918
DB 764 YTNDRCELIFTRAADIKGVANPQVSGVLGVWVPVGA--LKKMFALRLARPHQOMASVHQ 820
QY 819 DAMDSRVNPEGSFNSQFATKEEYTNVIANVVDKFSVSGITDPEMAPOVYSSTDGQF 878
DB 821 NAALDSRVNPEGSFNSQFATKEEYTNVIANVVDKFAEWGVTDPPEMAPOVYSSTDGSF 880
QY 879 LQSVIQNGVAFTRDYLKGSKANKYGTADQLVKAIKALHAKGLKVMADWPQOMYTFPKQ 938
DB 881 LQSVIQNGVAFTRDYLKGSKANKYGTADQLVKAIKALHAKGLKVMADWPQOMYAFPEK 940
QY 939 EVVTVTRTDFKGKPIAGSQINHSLYVTDTKSGDDYQAKYGGAFDELDELKEKPELFTKKQ 998
DB 941 EVVTVTRVKGYPVAVAGSQIKNTLYVVDGKSGKQQAQYGGAFLEELQAKYPELFAKQ 1000
QY 999 ISTGQALDPSVKIKQWAKYFNGSNILGRADYVLSQVSNKYFNVA--SDTLFLPSSILL 1056
DB 1001 ISTGVPMDPSVKIKQWAKYFNGSNILGRAGYVLSQVSNKYFNISDNKEINFLPKTLL 1060
QY 1057 GKVBSGIRYDGGYLYNSSATGDQVKASFIETAGNLYYFGKDGVMVTGAQTINGANYFF 1116
DB 1061 NQDSQVGFYDGGYLYYFT--SGYQAKNTFISEGDKWYFDNNGYVMTGAQSLNGVNYFF 1119
QY 1117 LENGTRALNTIYTDAGNSHYIYANDGKRYENGYYQF--GNDWRYFKDGNMAVGLTTVDGNV 1175
DB 1120 LSNGLQLRAILLKNEBDGTVAYGNDCRRYENGYYQFMSGVWRHFNNGEMSVGLTVIDGQV 1179
QY 1176 QYFDDKGVQAKDIIVTRDGKRYFPOHNGNAATNTFIADKTHWYLYLKGQVAVTGAQT 1235
DB 1180 QYFDEMGYQAKGFVTTADGKIRYFDKQSGNMYRNFIEENEGKWLVLGEDGAAVTGSQT 1239
QY 1236 VGKQKLYFRANGQOVKGDFTVTSDEGLYFYVDVDSGDMWTDFTIEDKAGNWFYLGKDAV 1295
DB 1240 INQCHLYFRANGVQVKGFEVTDHGRISYDNGSGDQIRNFRVNAQGVFFDNGYAV 1299
QY 1296 TGAQTIRGQKLYPKANGQOVKGDIVKGTGDKIRYDAXSGEQVFNKTVKAADGKTYVIGN 1355
DB 1300 TGAQTIRGQKLYFRANGVQVKGFEVTDYGRISYDNGSGDQIRNFRVNAQGVFFDNG 1359
QY 1356 DGVAVDPSVVGQTFKDAASGALRFPYMLKQGLVTGSGWYETANHDWYVIOGSKALTEQTI 1415
DB 1360 NGYAV-----TGARTI 1370
QY 1416 NGOHLYFKEDGHQVQLVTGDKRYRYDANSQDQAFNKSFTVNGKTYVYFGNDGTAQTA 1475
DB 1371 NGOHLYFRANGVQVKGFEVTDHGRISYDNGSGDQIRNFRV----- 1412
QY 1476 GNPKGQTFKDGSDIRFYSMEGQLVTGSGWYENAGQWLXY--KNGKVLTLGLTVGSRVYF 1534
DB 1413 -----RNAQGVFFDNGYAVTGAQTINGQHLXF 1442
QY 1535 DENGIOAKGKAVTSDGKIRYFDENGSGMI 1564

DB 1443 RANGVQVKGFEVTDYGRISYDANSGERV 1472
RESULT 13
AAU98035
ID AAU98035 standard; protein; 1475 AA.
XX AAU98035;
XX 27-AUG-2002 (first entry)
XX S. mutans glucosyltransferase GTFB mutant D457N/D571K.
KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture; mutant; mutein.
OS Streptococcus mutans.
OS Synthetic.
XX Key Location/Qualifiers
FH Misc-difference 457
FT /note= "Wild-type Asp substituted by Asn"
FT Misc-difference 571
FT /note= "Wild-type Asp substituted by Lys"
XX US2002031826-A1.
XX 14-MAR-2002.
XX 19-DEC-2000; 2000US-00740274.
XX 07-JUN-1995; 95US-00478704.
PR 07-JUN-1995; 95US-00482711.
PR 07-JUN-1995; 95US-00485243.
PR 16-JAN-1998; 98US-00007999.
PR 16-JAN-1998; 98US-00008172.
PR 20-JAN-1998; 98US-00009620.
PR 11-DEC-1998; 98US-00210361.
XX (NICH/) NICHOLS S E.
XX Nichols SE;
XX WPI; 2002-414332/44.
XX Glucosyltransferase B or D protein useful for producing a glucan useful
PT as substitutes for and additions to modified starch and latexes in paper
PT manufacture, comprises mutations in specific positions.
XX Claim 36; Page: 44pp; English.
XX The invention an isolated protein comprising a glucosyltransferase (GTF)
CC B polypeptide having changes at position from I448V, D457N, D567T,
CC K1014I, D457N/D567T, D457N/D571K, D567T/D571K, D567T/K1014I,
CC 1448V/D457N/D567T/D571K/K799Q/K1014I, Y169A/Y170A/Y171A, and K779Q or a
CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
CC complementary polynucleotide, a ribonucleic acid sequence encoding the
CC GTF mutant, an expression cassette comprising the polynucleotide operably
CC linked to a promoter, a vector comprising the expression cassette, host
CC cell introduced with the vector, a transgenic plant comprising the
CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
CC coating composition comprising a glucan produced in a plant transformed
CC with a gene encoding the mutant GTF, wild type or starch, a latex,
CC thermoplastic molecule or their combinations or glucan and starch where
CC the glucan is produced in the amyloplast and/or vacuole or a maize line
CC deficient in starch biosynthesis, transformed with a gene encoding a
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
CC comprising the glucan (paper sizing/coating agent). The vector is useful
CC for producing a glucan in a plant. The method comprises transforming a

amyloplast; vacuole; paper manufacture; mutant; mutein.
Streptococcus mutans.
Synthetic.

Key Location/Qualifiers
Misc-difference 457 /note= "Wild-type Asp substituted by Asn"
FT Misc-difference 567
FT /note= "Wild-type Asp substituted by Thr"

US2002031826-A1.

14-MAR-2002.

19-DEC-2000; 2000US-00740274.

07-JUN-1995; 95US-00478704.
07-JUN-1995; 95US-00482711.
07-JUN-1995; 95US-00485243.
16-JAN-1998; 98US-00007999.
16-JAN-1998; 98US-00008172.
20-JAN-1998; 98US-00009620.
11-DEC-1998; 98US-00210361.

(NICH/) NICHOLS S E.

Nichols SE;

WPI; 2002-414332/44.

Glucosyltransferase B or D protein useful for producing a glucan useful as substitutes for and additions to modified starch and latexes in paper manufacture, comprises mutations in specific positions.
Claim 36; Page; 44pp; English.

The invention an isolated protein comprising a glucosyltransferase (GTF) B polypeptide having changes at position from I448V, D457N, D567T, K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T, I448V/D457N/D567T/D571K/K1014T, Y169A/Y170A/Y171A, and K779Q or a GTF D polypeptide having changes at positions from T589D, T589E, N471D, N471D/T589D, and N471D/T589E. Also included are a glucan produced by the GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its complementary polynucleotide, a ribonucleic acid sequence encoding the GTF mutant, an expression cassette comprising the polynucleotide operably linked to a promoter, a vector comprising the expression cassette, host cell introduced with the vector, a transgenic plant comprising the vector, a seed or tuber from the transgenic plant, a paper sizing and/or coating composition comprising a glucan produced in a plant transformed with a gene encoding the mutant GTF, wild type or, starch, a latex, thermoplastic molecule or their combinations or glucan and starch where the glucan is produced in the amyloplast and/or vacuole or a maize line deficient in starch biosynthesis, transformed with a gene encoding a glucosyltransferase B or D enzyme, wild-type or mutant and a paper comprising the glucan (paper sizing/coating agent). The vector is useful for producing a glucan in a plant. The method comprises transforming a plant cell with the vector, growing the plant cell under plant growing conditions to produce a regenerated plant and inducing expression of the polynucleotide for a time sufficient to produce the glucan in the regenerated plant, where the vector contains a transit sequence from ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and chlorophyll AB binding protein to produce a transgenic plant, and glucan is produced in the amyloplast of potato or the vacuole of sugar beet. Glucans are useful as substitutes for and additions to modified starch and latexes in paper manufacture. Unlike prior art techniques, which require input materials that produce chemical effluents, paper manufacture utilising the glucan produced by GTF, which utilises biologically produced input materials, is more cost-effective and environmentally friendly. Moreover, glucans also exhibit thermoplastic properties and impart gloss to the paper during coating step. The present sequence represents a GTFB mutant of the invention. Note: The present sequence is not shown in the specification but was created by the indexer

Qy	939	EWTVTTRDKEGKPTAGSOINHSLVYTTDKSSGDDYQAKYCGAFDELKEKYPFLPYKKQ	998
Db	941	EVTATRTVDKVTGPVAGSQIKNTLYYVDVDSGSKDQQAQYCGAFLEELQAKYPFLPARKQ	1000
Qy	999	ISTGQAIDPSVKIKQWSAKYFNGSNILRGADYVLSDOVSNKYFNVA--SDTLELPSSLL	1056
Db	1001	ISTGVPMDPSVKIKQWSAKYFNGTNILRGAGYVLLKQATWYFNI SDNKEINFLPKTLL	1060
Qy	1057	GKVVESGIRYDGKGYIYNSSATGDQVKASFITEAGNLYYFKGQGMVTGAOTINGANYFF	1116
Db	1061	NQDSQVGSYDGKGYVYYST-SGYQAKNFTISEGDKWTYFDDNGYFMTGAOSINGVNYFF	1119
Qy	1117	LENGTALRNTIYTTDAQNSHYVANDGKRYENGYQOF-GNDWRYFKDGNNMAGLTVTDGVN	1175
Db	1120	LSNGLQLEDA LKNEDETGYAYGNDGRYENGYQFMSGWWHRFNNGEMSVGLTVIIDQV	1179
Qy	1176	QYFDKQGVQAKDKIIVTRDGKRVYFDQHNNGNAATNTFIADKTGHWYLYLGDGVAVTGAQT	1235
Db	1180	QYFDEMGYQAKGKFTVTTADGKIRYFDQSGNNWNRRIENEEGKWLVLGEDGAAVTGSQT	1239
Qy	1236	VGKQKLYFEANGQOVKGDFTVSDSKLYFYDVDVSGDMWTDFFIEDKAGNWFYLGKQGAAV	1295
Db	1240	INGQHLYFRANGVQVKGFEVTDHHRISYSDGNSGDQIRNRFVNAQGMWFYFDNNGYAV	1299
Qy	1296	TGAQTIRGQKLYFKANGQOVKGDIVKGTGDKIRYYDAKSGBOVFNKTKAADGKTYVIGN	1355
Db	1300	TGARTINGQLLYFRANGVQVKGFEVTDYBRYGRISYDGNSGDQIRNRFVRNAQGMWFYFDN	1359
Qy	1356	DGVAVDPSVVKGQTFDKDASGALRFYNLKGQLVGTSGWYETANHDDWYIYQSGKALTGBQTI	1415
Db	1360	NGYAV-----TGARTI	1370
Qy	1416	NGQHLYPEKDGHQVKGQLVTGTDCKRVRYSDANSGDQAFNKSVTYNGKTYFYFGNDGTAQTA	1475
Db	1371	NGQHLYFRANGVQVKGFEVTDHHRISYSDGNSGDQIRNRFV-----	1412
Qy	1476	GNPKGQTFKDGSDIRFYSMEGQLVTGSGWYENACQWLYY-KNGKVLITGLQTVGSGQRVYF	1534
Db	1413	-----RNAQGMWFYFDNNGYAVTGARTINGQHLYP	1442
Qy	1535	DENGIOAKGKAVRTSDGKIRYFDNDSGSMI	1564
Db	1443	RANGVQVKGFEVTDYBRYGRISYDANSGERV	1472

RESULT 15

AAU98036

ID AAU98036

XX

AC AAU98036; 322

XX
CVC-211X-LC
(F) 4040

DT 27-AUG-2002 (first entry)
XX

XX
DE S. mutants glucosyltransferase GTFB mutant D567T/D571K

XX
БФ
с: «Исание грусовуа

KW Glucosyltransferase;

KW coating composition;

KW amyloplast; vacuole;

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----

OS Streptococcus mutans.
OS Synthetic

US
XX
synthetic.

XX	FH	Key	Locat

FT Misc-difference 567

Query Match	54.7%	Score 4566	DB 5	Length 1475
-------------	-------	------------	------	-------------

Query Match: 34.7%; score 4300; DB 3
Best Local Similarity 56.2%; Pred. No 5e-268;

Matches 894; Conservative 212; Mismatches 340; Indels 144; Gaps 18;

.....

1 MEKNVRFKMHKVKRWVTLVSASATMLASALGASVASADTDTA---SDDSNQAVVTGDQ 56

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

b1 MDKVRYLKRVKKRWTVSVASAVMTLTTLGGGLVKADSNESKQISNDSNTSVVTANE 60

115 WVTENENQFTDEMLAEA-----KXVATAESDSIPSD-----LAKMSNUKQVDGKY 161
116 -----QNOTTVDKTSSEAAANNISKOTTEADTDVDDSSNAAMLQLEKLPNYKEIDGKY 169
162 YYYDQGNVKNFAVSGDKIYYFDETGAYKOTSKYDADKSSSAVSONATIPAAKNRAYS 221
170 YYYDNGKVRTNTLADGKILHFDGTATDTSIDTVNK--DIVTTRSILYKKNQOVD 227
222 TSAKNFEAVDNYLTADSWRPKSILKDGKTWTSBGKDDFRPLLMAMWPDTEKRNVTNYM 281
228 RSAQSFHDVHLYTAESWRPKYILKDGKTWTSBGKDDFRPLLMWPDQETQRQVNYM 287
282 NKVVGDIDTYAETSQADLTAAAEIVQARIEQKITSNNWTWMLREASAFVKTPQOWNGE 341
288 NAQLGINKTYDDTSNQLNLNAAATIAQIEAKITLTKNTDLWLRQTSIAFVKTSANSD 347
342 SERPYDDHLQNGALLFDNOTDLPTQSNYRLNRTPTNOTGSLDSRFTVNPNDPLGGYD 401
348 SERPFDHLQNGAVLDNEGKLTIPYANSYRLNRTPTNOTGSKDPRYT--ADNTIGGYE 405
402 FLAANDVDSNPVVQAEQLNWLHLYLNFSGSIYANDADANPDSIRVDAVDNVDADLLQIS 461
406 FLAANDVDSNPVVQAEQLNWLHFLMFGNIYANDPDANFDSIRVDAVDNVDADLLQIAG 465
462 DYKAAAYGIDKNNKNNANHVSIVEAWSDDNPTYLHDDGDMNMNDKFRLSMLWSLAKPL 521
466 DYKAAAGIHKNDKAANDHLSILEAWSDDNPTYLHDDGDMNMNDKRLSLFLSLAKPL 525
522 DKRSGNLPLHNSLDREVDTVPVPSYPARAHDSVODIIRIDIAEINPNPSFGYS 581
526 NQSGMNPILTNSLVNRTDNDATAVPSYSFIRAHDSVQTLIAKIIKAEINPNVVGYS 585
582 FTQEEIEQAFKIYNEDLKTKKYTHYNVPLSYTLTLTNKSGIPRVYVYGDMEFTDDGQYMA 641
586 FTWEEIKKAPETYNKDLATEKKYTHYNTALSYALLTLNKSVPVRYVYGDMEFTDDGQYMA 645
642 NKVTNTDAIESLLKARKMYVGGQAMQNYQIENGELITSVRYGKGALQSKDGKDATRTS 701
646 HKTINYEAITLLKARIKYVSGGQAMRNQOVGSEIITSVRYGKGALKATDGTDRTRTS 705
702 GVGVMGNQPNFSLDGK-VVALNMGAAHQAQVRAALMVSTKDGCVATYATDADAKAGLVK 760
706 GVAIVGNPESLKLKASDRVVMNGAAHKNQATRPLLITDNGIKAYHSDQEA--AGLVR 763
761 RTDENGLYFLNDLKGAVNPQVSGFLQVWPVGAADDQDIRVAASDTASTDGK--SLHQ 818
764 YTNDRGELIPTAADIKGYANPQVSGVLGVWPVGAA--LIKMFALRLARPHQOMASVHQ 820
819 DAAMDVRMPGFSNFSQSPATKEEYTNVVIANNVDKFSWGIITDFEMAPQYVSSTDGQF 878
821 NAALDSRVMPGFSNFSQAFATKKEEYTNVVIANNVDKFAEWGVTDFEMAPQYVSSTDGSF 880
879 LDSVIQNGVAFTRDYDLGMSKANKYGTADOLVKAIKALHAKGLKVMADWVPDQMYTTPKQ 938
881 LDSVIQNGVAFTRDYDLGSKPNKYGTADDLVKAIKALHSGKIKVMADWVPDQMYAFPEK 940
939 EVVTVTRTDKFGKPIAGSQINHSLYVYTDTKSSGDDYQAKYGGAFDLDELKYPFLFTKKQ 998
941 EVVTATRVDKYGTVPVAGSIKNLYLVYDGKSGKQQAQYGGAFLEELQAKYPELFARKQ 1000
999 ISTGQAI DSPVKIKQWSAKYFNGSNILGRGADYVLSQVSNKYFNVA--SDTLFLPSSLL 1056
1001 ISTGVPMDPSVKIKQWSAKYFNGTNILGRGAGYVLKDAQATNTYFNI SDNKEINPLPKTL 1060
1057 GKVVESIRYDGGKLYINSSATGDQVKASFIIEAGNLYYFGKDGVMYVTAQTINGANYFF 1116
1061 NQDSQVGSYDGGKGYTYIST-SGQAQNTFISEGDKWYFPDNNGYWVTAQSGINGVNYTF 1119
1117 LENGTAIRNTIYTDAGNSHYIYANDGKRYENGYQQF-GNDWRYFKDGNMAVGLTIVDGNV 1175
1120 LSNGLQLRDAILKNEIDGTAYYGNDRGRRYENGYQFMGSGVWRHFNNGEMSVGLTVIDGQV 1179

1176 QYFDKQGVQAKDKIIVTRDGKVRYPFOHNGNAATNTFIADKTGHWTYILGKDGVAVTGAOT 1235
1180 QYFDEMGVQAKGFVTTADGKIRYFDKQSGNMYRNFIEENEGKWLILGEDGAAVTGSQT 1239
1236 VGKQKLYFEANGQQVKGDFVTSDEGLXYFYDVSDDMMWTDFTIEDKAGNWFYLGKDGAAV 1295
1240 INGQHLVFRANGVQVKGFEVTDHGRISYYDGNSGDQIRNRFVRNAQGWFFYDNNGYAV 1299
1296 TGAQTIRGQKLYFKANGQQVKGDIYKGTGDKIRYDAKSGBQVFNKTVKAAQKTVVIGN 1355
1300 TGARTINGOLLYFRANGVQVKGFEVTDYGRISYYDGNSGDQIRNRFVRNAQGWFFYFDN 1359
1356 DGVAVDPSVVKGTQFKDASGALRFYNLKGQLVTGSGWYETANHDWVYIOSGKALTGEQTI 1415
1360 NGYAV-----TGARTI 1370
1416 NGQHLVFKBDGHQVKGQLVTGDKVRYYYDANDSGDQAFNKSVTVNGKTYVFGNDGTAQTA 1475
1371 NGQHLVFRANGVQVKGFEVTDHGRISYYDGNSGDQIRNRFV-----1412
1476 GNPKGOTFKDGSDIRPYSMEGQLVTGSGWYENAAQGWLTIV-KNGKVLTLGLQTVGSRVYF 1534
1413 -----RNAQGWFFYDNNGYAVTGARTINGOHLVYF 1442
1535 DENGIOAKGKAVRTSDGKIRYFEDENSGSMI 1564
1443 RANGVQVKGFEVTDYGRISYYDANDSGSERV 1472

Search completed: February 11, 2006, 19:14:59

Job time : 169.264 secs


```
Qy 402 FLANDVDSNPVQAEQALNHLNLFNFGSIYANDADANFDSIRVDVAVNDVADLLOISS 461
Db 406 FLANDVDSNPVQAEQALNHLNLFNFGSIYANDADANFDSIRVDVAVNDVADLLOIAG 465
Qy 462 DYLKAAAGIDKNNKANNHYSIYEAQNDNTPYLDHDDGNLMMNDKFRSLMSLAKPL 521
Db 466 DYLKAAAGIHKNDKAAADHLSILEASNDNTPYLDHDDGNLMMNDKFRSLMSLAKPL 525
Qy 522 DKESGLNPLIHNSLDREVDRVETVPSYSFARAHSEVQDIIRDIKAEINPNSPGYS 581
Db 526 NQBSGMPLITNSLVNRTDNDNAETAAVPSYSFIRAHSEVQDLIADIIKAEINPNVVGYS 585
Qy 582 FTQEEIEQAKIYNEDLKTKDKYTHYNVPLSYLLTNKGSIPRVYVGMFTDDGGYMA 641
Db 586 FTMEIEKKAPEIYNKOLLATEKKYTHYNVPLSYLLTNKGSIPRVYVGMFTDDGGYMA 645
Qy 642 NKTVNDYDAIESIAKARKYVSGGQAMQYQIENGELITSYRYGKALKQSDKGDATRTS 701
Db 646 HKTINYEAETLLKARIKYVSGGQAMRQVGNSEIITSYRYGKALKATDGTGDRTRTS 705
Qy 702 GUGVWVGNQNFSLDGK- VVALANGAHAHQEYRALMVSTKQGVATYATDADAKAGLVK 760
Db 706 GVAVIEGNNPSELKASDRVVMNGAAHKQATRPLLITDNGIKATHSDEA-AGLVR 763
Qy 761 RTDENGYLEFLNDDLKGVANPOVSGFLQVWVVPVGAADDQDIRVAASDTASTDGK--SLHQ 818
Db 764 YTNDRGELIFTAADIKGVANPOVSGYLGWVVPVGA--LHKMFALRLARPHQOMASVHQ 820
Qy 819 DAAMDVRMPEGSNFSQFATKBEETVNVVIANVVKFVSWGTDPEMAPQYVSSDGGP 878
Db 821 NAALDSRVMEGFSNFQAFATKBEETVNVVIAKNVDFKFAEWGVTDFEMAPQYVSSDGSF 880
Qy 879 LDSVIQNGVAFTRDYDLGMSKANKYGTADOLVKAIKALHAKGLKVWADVPDQMYTFPKQ 938
Db 881 LDSVIQNGVAFTRDYDLGMSKANKYGTADOLVKAIKALHAKGLKVWADVPDQMYTFPEK 940
Qy 939 EVVTVTRTDKFGKPIAGSQINHSLYVTDTKSSGDDYQAKYGGAFGLDELKEKYPELFTKKQ 998
Db 941 EVVTVTRDKYGTVPAGSQIKNTLYVVDGKSSGKQQAQYGGAFLEELQAKYFELFARKQ 1000
Qy 999 ISTGQALDPSVKIKQWSAKYFNGSNILGRGADYVLSQVSNKYFNVA--SDTLFLPSSL 1056
Db 1001 ISTGVPMDPSVKIKQWSAKYFNGSNILGRGAGYVLDQATNTYFNISDNKEINFLPKTL 1060
Qy 1057 GKVVESGIRVDGKGYLYNSSATGDQVAKAFITEAGNLVYFGKQGYMVTGAQTINGAYFF 1116
Db 1061 NODSQVGFSDGKGYVYIST-SGYQAKNTFISEGDKWYFPDNNGYMVTGAQSGINGVYFF 1119
Qy 1117 LENGTALRNTIYDAGNSHYANDGKRYENGYYQOF-GNDWRVFKDGNMAVGLTVDGNV 1175
Db 1120 LSNGLQLRALKNEDGTAYYNGDGRVYENGYYQPMGVRHFNNGESVGLTVIDGGV 1179
Qy 1176 QYFDDKGVQAKKIIIVTRDGKVRYPQHNNGNAATFIADKTHGWHYLLGKGVAVTGAQT 1235
Db 1180 QYFDENGYQAKGFVTTADKIRYFKQSGNMYRNFIEENEGKWLVLGEDGAATVGSQT 1239
Qy 1236 VQKQKLYFEANGQVKGDFVTSDEKLYFVDVDSGDMWTDFTIEDKAGNWFYLGKDGAAV 1295
Db 1240 INQGHLYFRANGVQVKGDFVTDHGRISYVDGNSGQIIRNFRVNAQOGWYFPDNNGYAV 1299
Qy 1296 TGAQTLRGKLYFKANGQVQKGDIVKGTGDKIRYDAKSGEQVFNKVAADGKTVYIGN 1355
Db 1300 TGARTINGQLLFRANGVQVKGDFVTDGRISYVDGNSGQIIRNFRVNAQOGWYFPDN 1359
Qy 1356 DGVAVDPSVVGQTFKXASGALRFYNLKGQLVTGSGWYETANHDWYVYIQSGKALIGEORT 1415
Db 1360 NGYAV-----TGARTI 1370
Qy 1416 NQGHLYFKEDGHQVQKQLVTGDKVRYVDANSQDQAFNKSVTVNGKTYVFGNDGTAQTA 1475
Db 1371 NQGHLYFRANGVQVKGDFVTDHGRISYVDGNSGQIIRNFRV----- 1412
```

```
Qy 1476 GNPKGQTFKDGSDIRFYSMEGQLVTGSGWYENAGQWLYV-KNGKVLTLGLQTVGSQRVYF 1534
Db 1413 -----RNAQGMFYFDNNGYAVTGARTINGHLYF 1442
Qy 1535 DENGIOAKGKAVRTSDGKIRYFDENSGSMI 1564
Db 1443 RANGVQVKGDFVTDGRYGRISYYDANSGERV 1472

RESULT 3
US-09-740-274-2
; Sequence 2, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott B.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-740-274-2
```

```
Query Match 54.9%; Score 4580; DB 2; Length 1475;
Best Local Similarity 56.4%; Pred. NO. 1.4e-313;
Matches 896; Conservative 212; Mismatches 338; Indels 144; Gaps 18;
```

```
Qy 1 MEKVRFPKHVKKRWVTLVSASATMLASALGASASADTDTA-----SDSNQAVVTGDQ 56
Db 1 MDKVRVYKLRKRWVTVSVASAVMTLTLLSGGLVKADSNBSKQISNDSNTSVVTANE 60
Qy 57 TTN--NQATDQTSIAATATSEQSASDTAATDQASAAEQGTGTASTDTAAQTNTNANEAK 114
Db 61 ESNVITEATSKQEAASSQNTNHTVTSSTSVVNPKE-----VVSNPYTVGETASNGEKL- 115
Qy 115 WYPTENENOGFTDEMLAEA-----KNVATASDSIPSD-----LAKMSNVKQVDGKY 161
Db 116 -----QNQTTVTDKTSSEAAANNISKQTTTADTDVDDSNAAQLILEKLPNVKEIDGKY 169
Qy 162 YYDQDGNVKKPAFVSGDKIYYFDBTGAYKOTSKVDADKSSASVQNAITPAANNRAYS 221
Db 170 YYDNNKGKVRTFTFIADGKILHFDETGAYTDTSDTVNKK--DIVTTRSNLYKKNQVYD 227
Qy 222 TSAKFEADVNYLTADSWYRPKSLKDGKTWTESGKDFRPLLMAMPDTETKRNVNYM 281
Db 228 RQAQSFVHDYLTAEWYRPKYILKDGKTWTQSTEKDFRPLLMAMPDTETKRNVNYM 287
Qy 282 NKWVGIDKTYTAAETSOADLTAALAAELVQARIQKITESENNTKWLREAI SAFVKTPQWNGE 341
Db 288 NAQLGINKTYDDTSNQQLNIAAATQAKIEATITLKNLWLRQITSAFVKTSQWNSD 347
Qy 342 SEKPVDDHLQNGALLFDNQTDLTPDTQSNRYLLNRTPTNQTGSLDSRFTYVNPNDPLGGYD 401
Db 348 SEKPFDDHLQNGAVLYDNEGKLTVPVANSYRILNRTPTNQTGKDPRT--ADNTTIGGYE 405
```

```
QY 402 FLLANDVDSNPVQAEQNLWHLNFGSIYANDADANFDSIRVDADVNDADLLOISS 461
Db 406 FLLANDVDSNPVQAEQNLWHLNFGNIYANDPDANFDSIRVDADVNDADLLOIAG 465
QY 462 DYLAAYAGIDKNNKANNNHVSIVEASDNDTPYLHDDGDNLMNDKNPRLSMLWSLAKPL 521
Db 466 DYLAAGKGIHKNDKAANDHLSLILEASDNDTPYLHDDGDNMNMNDKNLRLSLFSLAKPL 525
QY 522 DKRSGLNPLIHSNLDREVDDRETVPSYSFARAHSDSEVDIIRDIKAEINPNPSFGYS 581
Db 526 NQSRSGMPLITNSLVNRKTDNAETAAPVPSYFTRAHSDSEVDLIADIKAEINPNVVGYS 585
QY 582 FTQEEIEQAFKIYNEDLKTDDKXYTHYNVPLSYTILLTNKSGIPRVYVYGMDFDDCGQYMA 641
Db 586 FTWEEIKAFPEIYNKOLLATEKXYTHYNTALSVALLLTNKSSVPRVYVYGMDFDDCGQYMA 645
QY 642 NKTWYDAIESLKLARKMYVSGGOAMONTQIGNGEILTSVRYGKGALKOSDKGDATRTS 701
Db 646 HKTINTEALITLKAIRKYVSGGOAMRNOOVGNSEIITSVRYGKGALKATDGDRTRTS 705
QY 702 GGVVMGNOPNFSLDGK - VVALNMGAHANQETRALMVSTKDGVAITYATDADASKAGLVK 760
Db 706 GVAVIEGNNPFLKASDRVVMNGAAHKQAAYRPLLLTTDNGIKAYHSDQEA - AGLVR 763
QY 761 RTDENGILYFLNDDLGKVPANQVSGFLOVWPVPGAADDQDIRVAASDSTADGK - SLHQ 818
Db 764 YTNDRGELIPTAADIIRGYANPQVSGYLGWVWPVGAA - LKMFALRLARPHQOMASVHQ 820
QY 819 DAAMDVRMVEGFSNFQSFATKEEETNVVIANNVDFKFSWGITDFEMAPQVYSSTDGQF 878
Db 821 NAALDSVRMVEGFSNFQAFATKEEETNVVIAKNVDKFAEWGVTDFEMAPQVYSSTDGFS 880
QY 879 LDSVIQNGYAFTDRYDLGMSKANKYGTADQLVKAIKALHAKGLKVMADWVPDQMYTFPKQ 938
Db 881 LDSVIQNGYAFTDRYDLGISKPNKYGTADDLVKAIKALHSKGIKVMADWVPDQMYAFPEK 940
QY 939 EVTVTRTDKFGKPIAGSQINHSLYVTDTKSSGDDYQAKYGGAFIDELKEKYPELFTKKQ 998
Db 941 EVVTATRVDKYGTFAVAGSQIKNTLYVVDGKSGDQQAQYGGAFLEELQAKYPELFAKQ 1000
QY 999 ISTGQAIQPSVKIKQMSAKYFNGSNILGRGADVLSDOVSNKYFNVA - SDTLFLPSSL 1056
Db 1001 ISTGVPMDSVKIKQMSAKYFNGTNILGRGAGVYLDQDQATNTYFNISDNKEINFLPKILL 1060
QY 1057 GKVESGIRYDGGYIYNSSATGDQVKASFITEAGNLYYFGKDGVMYVGTQATINGANYFF 1116
Db 1061 NQDSQVGFSDYDKGVYVYST - SGYQAKNTFISEGDKWYVFDNNGYVMTGAQSGINGVNYF 1119
QY 1117 LENGTALRNTIYDAQGNSHYYANDGKRYENGYYQOP - GNDWRYFKDGNMAVGLTVDGNY 1175
Db 1120 LSNGLQLRDAILKNEDGTAYAYGNDGRRYENGYYQFMSGVYMRHFNNGEMSVGLTVIDGQV 1179
QY 1176 QYFDKQGVQAKDIIIVTRDGKRYVFOHNGNAATNTFIADKTHWYVYLGKDGVAVATGAOT 1235
Db 1180 QYFDEMGYQAKGFVTTADGKIIRYFKQSGMYMRNFIENECKNLYLGEDGAATVGSQT 1239
QY 1236 VGRQKLYFEANGQOVKGFVTSDEGLKLYFYDVSQDMWTDFTIEDKAGNWFYLGKDAV 1295
Db 1240 INGCHLYFRANGVQVKGFEVTDHGRISYYDNGSGDQIRNRNFRVNAQOGQWFFDNGYAV 1299
QY 1296 TGAQTIRGKLYFKANGQOVKGDIVKGTDKIRYBYDAKSGEQVFNKTVKAADGKTYVIGN 1355
Db 1300 TGARTINGOLLYFRANGVQVKGFEVTDYGRISYYDNGSGDQIRNRNFRVNAQOGQWFFDN 1359
QY 1356 DGVAVDPSVVGQTFFKXASCALRFPYNLKGQVLTGSGWYETANHDWVYIQSGKALTGEQTI 1415
Db 1360 NGYAV - - - - -TGARTI 1370
QY 1416 NGQHLFYKEDGHQVKGQVLTGTDGKRYVYDANSGDQAFNKSVTVNGKTYFNGDGTQATA 1475
Db 1371 NGQHLFYFRANGVQVKGFEVTDHGRISYYDNGSGDQIRNRFV - - - - - 1412
QY 1476 GNPFGQTFKDGSDIRFYSMEGQVLTGSGWYENAGQWLYV - KNGKVLITGLQVTSQRYVF 1534
```

```
Db 1413 -----RQAQGWFFDNGNGYAVTGARTINGOHLIF 1442
QY 1535 DENGIAQKAVRTSDGKIRYFDENSGSMI 1564
Db 1443 RANGVQVKGFEVTDYGRISYYDANSGERV 1472

RESULT 4
US-09-210-361-4
; Sequence 4, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-210-361-4

Query Match 50.5%; Score 4214.5; DB 2; Length 1375;
Best Local Similarity 60.3%; Pred. No. 7.4e-288;
Matches 811; Conservative 186; Mismatches 296; Indels 53; Gaps 14;

QY 1 MEKNVRFPKHVKRKRWVTLVSASATMASALGASVADTDASDSDNQAVVTGDDT--T 58
Db 1 MEKKVRFLKRVKRWVTVSIASAVVTLTSLGSLVKAD---STDDRQQAUTESQASLVT 57
QY 59 NNOATDOTSIA--ATATSEQSASTDAATDQASAAEQTOGTAST--DTAAQTITNANEA- 113
Db 58 TSEAAKETLTATDTSTATSATSQPTATVTDNVSTTNQSTNTTANTANFVVKPTTSEQAK 117
QY 114 -----KWVPTENEN---QGFTDEM-----LAERAKNVATAESDS 143
Db 118 TONSDKIITTSKAVNRLTATGKFPVANNNTAHPKTVTKIVPIKPKIGLKQPSLSQDD 177
QY 144 IPSDLAKMNVQVDGKYVYVYDQGNVKNFAVSVDGKIYYDETCAYKDTSKVDADKSS 203
Db 178 IAA-LGNVNIIRKRVNGKYVYKEDGTLQKNYALNNGKTFEFDGTGALSNNLT-LPSKCKGN 235
QY 204 SAVSONATIFAAANNRAYSTSAKNFEADVNLVADSWYRPKSILKDGKTWTESGKDDFRPL 263
Db 236 ITNNDNTNSFAQYNQVYSTDVANFEVHDYLTAEWSYRKYILKDKGKTWTQSTEKDFRPL 295
QY 264 LMAWPDPTKRYNVNMKNVVGIDKTYTAETTSQADLTAAAEVLQARIKQITSENNTKW 323
Db 296 LMTWPDQETQRQYVNMNAQLGIHQYNTATSPQLNLAAQTIQTKIEKITAENKTNW 355
QY 324 LREAISAFVKTQDQWNGESEKPYDDHLQNGALLFDNQDTLTPDTQSNYRLNPTNQTG 383
Db 356 LRQTISAFVKTQSAWNSDEKPFDDHLQGALLYSNNSKLTQSAANSYRLNPTNQTG 415
QY 384 SLDSRFTYPNPDPLGGYDFLLANDVDNSPNVQAEQNLWHLNFGSIYANDADANFDS 443
```


Qy	1093	LYFEKDCGYMTGAQTINGANYFFLENGTALRNTIYTDAQNSHYIYANDGKRYENG'QOF	1152
Db	1127	WYFDNNHMYGLQQLNGEVOYFLSNGVQLRESFLENADGSKNYFGHLGNRYNGSYSF	1186
Qy	1153	GND--WRYF-KDGNMAVGLTTVDGNVOYFDKDGVOAKDKIIVTRDGVKRVYFDQHNGNAAT	1209
Db	1187	DNDSKWRYPDASGMAVGLKTINGNTQYFDQDGYQVKGAWITSGDGKKRFPDGSQNMV	1246
Qy	1210	NTFIADKTGHYYILKGQSVAVTGAQTGVGKQKLYFEANGQQVKGDFVTSDEGLYFYDVS	1269
Db	1247	NRFANDKNGDWYILNSDGIALVQQTINGKTYYPGQDGKQIKGKIIT-DNGKLYFLANS	1305
Qy	1270	GDMWTDPIEDKAGNWFYLGKDGAAVTCQITRQKLYFRANGQOVKGDIVKGTDKIRY	1329
Db	1306	GELARNIPATDSQNNWYIFGSDGVAVTGSQTIAGKLYFASDGGQVKSFT-YNGKVHY	1364
Qy	1330	YDAKSGEOVFNKTKAADGKTYIVTNGDVAVDPSVWKQT'FKDASGALRFVNLKGQLVTG	1389
Db	1365	YHAUSGELQVNRFEADKDG-----	1383
Qy	1390	SGWYETANHDMVYI'QS-GKALTGSQTINGQHLYPKEDGHQVKGOL	1433
Db	1384	-----NWYILDSNGEALTQSORINDORVFFTFREGKOVKGDV	1419

```

RESULT 7
US-09-210-361-6
; Sequence 6, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
; US-09-210-361-6

```

```

Query Match      ! 42.8%; Score 3575; DB 2; Length 1430;
Best Local Similarity 48.4%; Pred. No. 9.1e-243;
Matches 719; Conservative 240; Mismatches 408; Indels 118; Gaps 26;

Qy 1 MEKNVRKMKHKVKRWVTLVSVA--TWLASALCASVAS-----ADDTAS 44
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 METKRRYKMKHKVKRWVTVASGLITLGTTLGSSVSAETEQTSDKVKVTQKSEDDKAA 60
      |||:|||||:|||||:|||||:|||||:|||||:|||||:

Qy 45 DDSNQA---VVTGDTTNNQATDTSIAATATS-EQSASTDAAATDQASAAEQTGTTAST 100
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 SESSQTDAPKTKQAQTEQTQAQSANVADTSITKETPQNITTOANSDDKTVNTKSE 120
      |||:|||||:|||||:|||||:|||||:|||||:|||||:

Qy 101 D-TAAQTNTNNAEKVPTPENENQGFTEMLAEAKNVAT-AESDSIPSLAKNKNVQKV 157
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 EAQTSERTKQSEBAQ---TTASSQALTKQAELTKQRTAAOSKNQKPVDLAAATPNVKQI 177
      |||:|||||:|||||:|||||:|||||:|||||:|||||:

```

Qy	158	DKKYYDDQGNVKNFPAUSVGDKIYYFDE--TGAYKDTSKVDADKSSSAVSONATIPFAAN	211
Db	178	DKKYYGSDGQPKNFALTYNNKVLYPDKNKTGALTDTSTQYQFKQGLTKLND---YTPH	234
Qy	217	NRAYSTSAKNFEADVNYLTADSWTRPKSILKDGKTKWTESGKDDRPRLMLAMWPDDETGRN	276
Db	235	NQIVNFENTSLETTDNYVTADSWTRPKDILKNGKTKWTASSESDLRPLMSWMPQKQTQIA	294
Qy	277	YVYNNKV--VGIDKITYTAETSSQADLTAAAEVLQARIEQKITSENNTKWLREBAISAFVKQT	335
Db	295	YLYNMNQGLGTGENYTADSSQESINLAAQTVQVKIETKISQTOQTQOTQWLDRDINSFVKQT	354
Qy	336	PQWNGESEKPYD----DHLQNGALLFDNOTDLTPDTOSNYLLNRTPTNQTSLSDRFTY	391
Db	355	PNWNSOTSEDTSSAGEKHOHLOGGALLYSN--SDKTAYANSDYVRELLNRTPTTSQTG----	409
Qy	392	NPNDPLGCGYDFLLANDVDSNPVQAOSQLNWLHYLLNFGSTYANDADANFDSIRVDADV	451
Db	410	FEDNSGGYDFLLANDIDNSNPVQAOSQLNWLHYLLMYGSTIVANDPEANFDGVRVDADV	469
Qy	452	VDADLLQTSDDLKAAAGIDKNNKNANHHVSIVEAWSNDPTPYLHDDGDNLMNMNDPRL	511
Db	470	VNADLLQIASDYLKAHYGVDKSEKNAINHLSILEAWSNDPQYNKDTKGAQLPDKNLRL	529
Qy	512	SMLWSLAKPLDK-----RSLNPLIHNSLDVREVDREVETVPYSFAPAHDSVQD	563
Db	530	SLLYALTRPLEKDAKSNKNEIRSGLEPYITNSLNRSAEGKNSRMANYPITRAHDSVQT	589
Qy	564	IIRDIKAEINPNSFGYSFTOEETEOQAFKLYNEDLKKTDKKYTHYNVPLSYTLLTNKGS	623
Db	590	VIAKIIKAQINPKTDGUTFTLDELUKAPKIYNEDMRQAKKKTOSNIPTAYALMSLNKDS	649
Qy	624	IPRVYGDMMFTDQOYMANKTVNYDAIESLLKARMKTVSGGQAMQ--NYQIENGSE----	676
Db	650	ITRLYGDMSDDGOYMATKSPYDAIDTLLKARIKYAAGQDMKIIYVEGDGSHMDWDY	709
Qy	677	--ILTSVRYGKALKQSDKGDAITRTTSGVGVVMGNQPNFSLD--GKVVALNMGAAHANQ	733
Db	710	TGVLTSVRYGTGANEATDQGEATKTQCMAVITSNPNLKLNQNDKIYVNMGAHKNQ	769
Qy	734	RALMVSTKDGVAITYATDADASKAGLVKRTDBENGVLIFLNDLKGVAPOVSGFLOVWPV	793
Db	770	RPLLTAKTKDGLTSYTSDAAK--SLYKTKNDKGELVFEDASDIQYLYNPQVSGYLAVWPV	827
Qy	794	GAADDQIRVAASDTASTDGLSLHODAAAMPVYMFEGFSNFQSPATKEEYETNVVIANNV	853
Db	828	GASNDQVRVAASNKANATQVYESSSALUSQLIYEGFSNFPDFTVKDSYTNKKIAQN	887
Qy	854	DKFVSGITDFEMAPQYVSSTDGQFLDSVIQNGYAFTRDYLGLMSKANKTGTADQLKAI	913
Db	888	QLPKSGVTSFEMAPQYVSSSEDGSFLDSIIQNGYAFEDRYDLAMSKNKYGSQODMINAV	947
Qy	914	KALHAKGLKMWADWPDMQYTPPKQEVVTVTRTDKFKGPIAGSOINHSLYVTDTKSSGDD	973
Db	948	KALHKSIGIADWVPDQIYNLPECKEVYVTRVNDYGEYKDSSEIKNTLYAANTKSNKG	1007
Qy	974	YQAKYGAFPLDELKEKYPELFTKKQJSTGQAIIPSVKIKOWSAKYFNGSNITLGRGADYVL	1033
Db	1008	YQAKYGAFSELAAKYPSIFNRTQISNGKKIIPSEKITAWKAKYFNGTILGRGVGYVL	1067
Qy	1034	SDOVSNKYFNVAADTLPLPSLLGKVVESGIRYDGKGYIYNSSATGDQVKAISFITEA--GN	1092
Db	1068	KDNASDKYPELKGNOTYLPKQMTNKEASTGFVNDGNGMTFYST--SGYQAKNSFVQDAKGN	1126
Qy	1093	LYTFGKDGVMYVTAQTINGANYFFLENGTALRNTIYTDAGNSHHYANDGKRVYENGQQF	1152
Db	1127	WYTFPDNNGHMVYGLQQLNGEVQYFSLSGVQLRESFLENADGSKNYFCHLGNRYNSGYYSF	1186
Qy	1153	GND--WRYF--KDCGNMAVGLTVDGNVOYFDKCGVQAKKIIIVTRDGVKRYFDQHNGAAT	1209
Db	1187	DNDSKMYIFDASGVMAVGLKTTINGNTQYFPDQDGYQVKGAWITGSDGKKRYFDDGSGMNAV	1246
Qy	1210	NTFTADTKGHYYLKGDKGVAVTGAQTVGKQKLYFEANGQOVKGDFVTSDESKLYFYDVDS	1269

Db 1247 NRPANDKNGDWYLYNSDGLALGVQTINGKTYFFQDQKQIKGKIIT-DNGKLKYP LANS 1305
QY 1270 GDMWTDFTEDKAGNFWYLGKDGAAVTGAQTIRGOKLYFKANGQQVKGIVRGTDGKIRY 1329
Db 1306 GELARNIFATDSQNNWYFGSDGAVTGSQTAGKLYFASDGKQVKSFTV-YNGKVHY 1364
QY 1330 YDAKSGEQVFNKTVKAADGKTYVYVGNQVAVDPSPVVGKQTFKDGASGALRFPYNLKGOLVTG 1389
Db 1365 YHADSGLQVNRFEADKDG-----1383
QY 1390 SGWYETANHDWYIOS-KHALTCEQTINGOHLYFKEDGHQVKGOL 1433
Db 1384 -----NWYLYDSNGEALTGSORINDQRFVFTREGQVKGDV 1419

RESULT 8

US-09-740-274-6
; Sequence 6, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740, 274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-6

Query Match 42.8%; Score 3575; DB 2; Length 1430;
Best Local Similarity 48.4%; Pred. No. 9.1e-243;
Matches 719; Conservative 240; Mismatches 408; Indels 118; Gaps 26;

QY 1 MEKNVFPKHKVKKRWVTVLSVASA--TMLASALGASVAS-----ADTDITAS 44
Db 1 METKRYKHKVKKRWVTVAVASGLITLTGTTLLGSSVSASAEQTSKVKVTKSEDDKAA 60
QY 45 DSDNQA---VVTGDTNNQATDQTSIAATFATS-EQASASTDAATDAQASAEQQTGTTAST 100
Db 61 SESSQTDAPKTKQAQTEQTAQSQANVADTSTSIKTPSQNTTQANSDDKTVTNKSE 120
QY 101 D--TAQOTTNANEAKWVPTENENQGFTEMLAEAKNVAT-ABSDSIPSDLAKMSNVQV 157
Db 121 EAQTSBERTKQSEBAQ---TTASSQALTOAKAELTKQRTAAQENKNPVDLAAIPNVQI 177
QY 158 DGKYYVYDDGNNKFNFAVSGPKIYYEDB-TCAYKDTSKVDADKSSASVQNAATIPAA 216
Db 178 DGKYYVYIGSDGQPKKQNFALTVNNKVLVFDKNTGALTDTISQYQKGLTKLND---YTH 234
QY 217 NRAYSTSAKNFEAVDNYLTADSWYRPRKSLKDGKWTWESGKDFRPLLMWVPDPTKRN 276
Db 235 NQIVNFENTSLEITIDNVYVTDADSWYRPRKDLKNGKWTWASSESDLRPLLMWVPDPTKRN 294
QY 277 YVNMKV-VGIDKTYTAETSQADLTAAAEVLQARTEOKITSENNTKWLREALSAFVK 335

Db 295 YLYNWNQOGLGTGENYTDASSQESLNLAAQTQVQVKIETKISQTOQOWLRDINSEFVK 354
QY 336 POWNGESEKPYD----DHLQNGALLFDNQTDLTPTQSNRYLLNRTPTNQTSLSRFTY 391
Db 355 PNWNSQTESDTSAGEKQHLQGGALLYSN-SDKTAVANSDYRLNRTPTTSQTG---KPKY 409
QY 392 NPNDPLGGYDFLLANDVNSNPVQABOLNWLHYLLNFGSIYANDADANFDSIRVDADVN 451
Db 410 FEDNSSGGYDFLLANDIDNSNPVQAEQLNWLHYLLNFGSIYANDPEANFDFGVRVDVN 469
QY 452 VDADILQJSSDYLKAAYGIDKNNKNNHVSIVAEASDNDTPYLHDDGDNLNMDNKPFL 511
Db 470 VNADILQJASDYLKAHYGVDSKNAIHLNLSLEAWSNDNDPOYNKDTKGAQAPIDNKLRL 529
QY 512 SMLSLAKPLDK-----RSLNPLIHNSLVREDDRETVETVPSYSFARAHSEVOD 563
Db 530 SLLYALTRPLEKDAASKNEIRSGLEFPVITNSLNNSAEGKNSERMANYIFIRAHSEVQT 589
QY 564 IIRDIKAEINPNSFGYFTQBEIEQAFKIYNEDLKKTKDKKTHYVNPUSLYTLLTNKGS 623
Db 590 VIAKIKAQINPKTDGLTFLDELQAKFIYNEDMROAKCKYTQSNIPITAYALMSNKDS 649
QY 624 IPRVYVYGMFTDDGQYMAKNTVYDAIESLLKARKKYVSGGOAMO-NYQINGE----- 676
Db 650 ITRLYIGDMYSDGQYMATKSPYDAIDTLLKARIKIAAGQDMKITYYEGDKSHMDWDY 709
QY 677 --ILTSVRYGKALKQSDKGDATRTTSGVVMGNQPNPFLD-GKVVALNMGAHAHQY 733
Db 710 TGVLTSVRYGTGANEATDQGSSEATKTQGMVAVITSNPNSLKLQNDKVIYVMGAHKNQY 769
QY 734 RALMVSTKDGAVATYATDADAKAGLVKRTDENGNYLYFLNDDLKGVANPOVSGFLQVWVPV 793
Db 770 RPLLLTTKDLGTSYTSDAAK--SLYRKTNDRGELVFDASDIQGYLNPQVSGYLAVMVPV 827
QY 794 GAADDODJRVAASTDASTDGKSLHODAAAMDSRVMEFGFSNFOSFATKBEYTNVTVANNV 853
Db 828 GASDNDQVRVAASNKANATQVYESSALDSQLYTEGFSNFQDFTVKDSYTNKKIQAQNV 887
QY 854 DXFVSWGITDFEMAPQYVSSSTDQQLDSVIONGYAFTDRYDLGMSKANKYGTADQLVKA 913
Db 888 QLFKSGVTSFEMAPQYVSSSEDSFLDSIIQNGYAFEDRYDLAMSNNKKNYKGSQODMINAV 947
QY 914 KALHAKGLKVMADWPDQMYTTPKQEVTVTTRDFKPGKPIAGSQINHSIYVTDTKSSGDD 973
Db 948 KALHKSQIQVIADWVPDQIYNLPKQEVTVATRVNDYGEYRKDSEIKNTLYAANTKNGKD 1007
QY 974 YQAKYGGAFDELKEKYPELFTKKOISTGOATDPSVKIKQWSAKYFNGSNILGRGADYVL 1033
Db 1008 YQAKYGGAFLELAAYKIPSFNRTQISNGKKIDPSEKITAWAKAYFNGNILGRGVYVL 1067
QY 1034 SDQVSNKYFNVASDTLFLPSSLLGKVVESGIRYDGKGYIYNSSATGDQVQKASPIEA-GN 1092
Db 1068 KDNASDKYPELKGNTYLPKQMTNKEASTGFVNDGNGMTFYST-SGYQAKNSFPVQAKGN 1126
QY 1093 LYYPKDGKGMVTAQTINGANFYFLENGTALRNTIYTDAGNSHHYIYANDKRYENGYYQOF 1152
Db 1127 WYTFDNNGHMYGLOQLNGEVQYFVLSNGVQLRESFLENADGSKNYFCHLGNRYSNGYYSF 1186
QY 1153 GND--WRYF-KDGNNAVLGTLTVDGNVQYFDKGVQAKDKIIVTRDCKVRYFDQHNGNAAT 1209
Db 1187 DNDKRWYFADSGVMVAVGLTKNTGNTQYFDQDGYQVKGAWITGSDGKKYFDDGSGNMAV 1246
QY 1210 NTFIADKTHWYTLGKDGVAVTGAQTQVGVQKLYFEANGQQVKGDFVTSDEGLYFYDVDS 1269
Db 1247 NRPANDKNGDWYLYNSDGLALGVQTINGKTYFFQDQKQIKGKIIT-DNGKLKYP LANS 1305
QY 1270 GDMWTDFTEDKAGNFWYLGKDGAAVTGAQTIRGOKLYFKANGQQVKGIVRGTDGKIRY 1329
Db 1306 GELARNIFATDSQNNWYFGSDGAVTGSQTAGKLYFASDGKQVKSFTV-YNGKVHY 1364
QY 1330 YDAKSGEQVFNKTVKAADGKTYVYVGNQVAVDPSPVVGKQTFKDGASGALRFPYNLKGOLVTG 1389

Db 1301 DKFIRADGTRFYIPDTGNLAVNRFAQNPENKAWYYLDSNGYAVTGLQTINKQYYPDN 1360
Qy 1246 NGQOVKGFVTSDEKLYFYVDVDSGDMWTDFTFIEDKAGNWFYLGKDGAAVTAQTIRGQK 1305
Db 1361 EGRQVKGHFVITNNQR-YFLDGSGBIAPSRFVTEN-NKMYYYDGNGLVKVGAQVINGNH 1418
Qy 1306 LYFKANGQOVKGDIVKGTGDKIRYDADKSGEQVFNKTVRAADGKTYVIGNDGVADPVSU 1365
Db 1419 YYPNNDYSQVKGAWANG-----RYDGDGSAQVSNQFI----- 1451
Qy 1366 KGQTFKADSGALRFYNLKGQVLTGSGWYETANHDWYI-QSGKALTEQOTINGHLYFKE 1424
Db 1452 -----QIAANQWYLAQDGHKVTGLQINNNKVVYFGS 1483
Qy 1425 DGHQVKQLVTGDKVRYVDANSQDAFNKSVTVNGKTYVYFNGDGTAGTAPNKGQTFK 1484
Db 1484 NGAQVKGKLLT-VQGKKCYFDAHTGEQVNRV----- 1515
Qy 1485 DGSDFRYSMEGQLVTGSGHYENAGOWLYVKN-GKVLTLQTVGSQRVYFDENGIOAKG 1543
Db 1516 -----EAARGCWYFNSAGQAVTGGQVINGKQLYFDGSGRQVKG 1554
Qy 1544 KAVRTDQKIRYFDENSGSM 1563
Db 1555 RYVYVG-GKELFCDAKTGEL 1573

RESULT 10
US-09-499-203-2
; Sequence 2, Application US/09499203
; Patent No. 6570065
; GENERAL INFORMATION:
; APPLICANT: KOSSMANN, Jens
; APPLICANT: WELSH, Thomas
; APPLICANT: QUANZ, Martin
; APPLICANT: KNUTH, Karola
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternanase
; FILE REFERENCE: 147-196P
; CURRENT APPLICATION NUMBER: US/09/499, 203
; CURRENT FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2057
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-09-499-203-2

Query Match 29.4%; Score 2454; DB 2; Length 2057;
Best Local Similarity 37.4%; Pred. No. 1.6e-163;
Matches 627; Conservative 235; Mismatches 541; Indels 274; Gaps 58;

Qy 35 VASADTDTASDSDSQAVVTGDTT--NNQATDQTSIAATATSQSASTDAATDQASAAEQ 92
Db 242 VADSSGQTYFDNGQPLGLQLIIDLGNLQVFNQGVQI-----KGSFQDVNNKRIYFAPN 236
Qy 93 TQGTASTD-----TAAQTITTNAEAKWVPPTENENQGFDTBMLAEAKNVAFAESDPSDL 148
Db 297 TGNVAANTEIINGKLQGRDANGQVK-----NAFSKDV---AGNTFYFDANGV---- 341
Qy 149 AKMSNYKQVDGKYYDYDDGDNVKNFAVSGDKIYFD-ETGAYKDTSKVDADKSSAVS 207
Db 342 -MLTGLQTSIGKTYLDEQGHRLKRYAGTFNNQFMFYFDADTGAGKTAIEYQFPQGLVSQS 400
Qy 208 QNATIFAANNRAYTSKAKNFEADVNYLTADSWRPKSIKDGKTWTESGDKDPRPLLMW 267
Db 401 NENT---PHNAKSYDKSFENVGVYLTADTWRTPTDLKNGDTWTASTETDNRPLMTW 457
Qy 268 WPDTEKRNYYNM-NKRVGIDKTYTAETSQADLTAAAEVLQARIQKITSNNKWLRE 326
Db 458 WPDQIQOANYLNFMSKGLGITTTYYTAATSQKTLNDAAFVIQTAIEQIISLKSKSTEWLRD 517
Qy 327 AISAFAVKTQPOWNGESE-KPYD--DHLQCALLFDNQDTLTPDTQS-NYELLNRTPTNQT 382

Db 518 AIDSPVKTOANWKNQTEDEAFGLQWQGFPLAYQDSSHRTPTNTDSSGNRKLGQRQINID 577
Qy 383 GSLDSRFTYPNPDPLGGYDFLLANDVDSNPVQVQAEQLNMLHYLLNFGSIYANDADANFD 442
Db 578 GSKDT-----TDGKGS-EFLLANDIDNSNPVQAEQLNMLHYLMNFGSITGNDNANFD 630
Qy 443 SIRVDAVDNVDADLLQISSDYLKAAAGIDKNNKNANNHVSIVEAWSNDNTPYLHDDGDL 502
Db 631 GIRVDAVDNVDADLLKIAAGDYFKALYGTDKSDANANKHLISILEDWMGKDPQVYVNOQNAQ 690
Qy 503 MNMKNKFRLSMLWSLAKPLDKRSG-----LNPLI-----HNSLVDREV 541
Db 691 LTWDYTVTSQFNSLTHGANNRSNMWYFLDTGYLLNGDLNKKLVKDNKRPNSTGLVNRAN 750
Qy 542 DREVETVPSYSPARAHSEVDQIIRDIKAEINP---NSFGYSFTQEEIEQAFKIYNEDL 598
Db 751 SGTQKVPINYSFVRAHDYDAQPIR---KAMIDHGIKMMQDTFTFDQLAQGMFYYKQD 807
Qy 599 KKTQ--KKYTHNVPLSYTLTLTNKGSIPRVYVGMFTDDGQYMANKTNYDAIESLLKA 656
Db 808 ENPSGPKKYNDYNLPSAYAMLLTNKDTVPRVYVGMYLEGGQYMEKGTIYNPVISALLKA 867
Qy 657 RMKYVSGQAM-----QNYQIENGELTSLVRYGKALKQSKDGDATT-----RTSG 702
Db 868 RIKYVSGGQTMATDSSGKDLKGETDLLTSVAFKGIM---TSDQTTTQDNSQDYKNOG 923
Qy 703 VGVVMGNQPNFSLDG-KVVALNMGAAHANQRYALMVSTKDGVAITYATDADAKAGLVR 761
Db 924 IGVIVGNPDKLNNDKTITLHMKAHKNQLYRALVLSNDSGIDVDYDSDKAP---TLR 979
Qy 762 TBENGILYF-----LNDDLKGVANPQVSGFQVWVVPVCAADDQDIR-VAASD 807
Db 980 TNDNGDLIFHKNTTFVKQDGTIINYEMKGLNALISGYLGVWVVPVGSASQDARTVATES 1039
Qy 808 TASTDGSLSLHODAAAMDSTRVMEFGSNFQSFATKEBYTNVANNVANNVDFKVSMTITDFEMA 867
Db 1040 SSSNGSVFHSNAALDSNVIYEGFSNFQAMPTSPQSTNVVIATKANLKFELGITSFELA 1099
Qy 868 PQVYSSTDG-----QFLDSVIONQYAFTRDYDLGMSKAN-----KYGTADQLVKAIAL 916
Db 1100 PQYRSSGDTNYGMSFLDSFLNNGYAFTDRYDLGFNKGADGNPNPTKYGTQDQLRNLAEAL 1159
Qy 917 HAKGLKWMADVPDQMYTFPKQEVTVTRTFKFKPIAGSQINHSIYVTDTKSSGDDYQA 976
Db 1160 HKNGMQAIADWVPDQIYALPKGEVTVATRVDERGNQLKDTDFVNLLYVANTKSSGVDYQA 1219
Qy 977 KYGGAFLDELKPKYPELFTKKQISTGOAIDPSVKIKOWSAKYPNGSNILGRGADYVLSQ 1036
Db 1220 KYGGEFLDKRBEYPSLFRQNGVSTGQPIDASTKIKOWSAKYNNGNINILHRGAYYVLKDW 1279
Qy 1037 VSNKYFNVA-SDTFLPSSLLGKWBESGIRYDGKGIYINSSATGDQVKASFITEA-GNLY 1094
Db 1280 ATQYFNIAKTNEVFLPLQLQNKDAQTFISDASGVKY-YSIGYQAKDTFIEDGNGWY 1338
Qy 1095 YFGKQGMV-----TGAQTINGANYFFLENGTALANTYTDAGQNSHYIAND 1141
Db 1339 YFDKQGMVRSQOGENPIRTVETSVNTRNG-NYFMPNGVELRKFGFTDNGSNVYFPDDQ 1397
Qy 1142 GKRENGY--QQFGNDWRYFKDGNMAVGLTTVDGN-VQYFDKDGVAQAKDKIIVTRDGKVR 1198
Db 1398 GKNVRDKYINDDANNFYHLNDGTVMSRGLFKFSDSDTLQIFASNGVQIKDSYAKDSKGNKY 1457
Qy 1199 YFDHNGNAATFTIADKTGHW-----YLLGKDGVAVTGAQTGVKQKLY--FEANGQOVKG 1252
Db 1458 YFDSATGNNDTG-----KAQTDWNGYIYITSDA---NNTIGVNTDYATYITSSLEDG 1509
Qy 1253 DFVTSDEGLKLYFVDVDSGDMWTDFTFIEDKAGN---WFLYLGKOGAAVTGAQTIRGQKLYFK 1309
Db 1510 LFANAEPYG-----VVTQDQNGDLKMQY-----INHTKQY-- 1539
Qy 1310 ANGQOVK-----GDIVKGTGDKIRYDAKSQGEQVFNKTVKAADGK 1349

Db 1540 -EGQOVQVTRQYTSKGVSNWNLITPAGGDL-----QGRLWVDSRALJWTPFKTNQISFI 1594
Qy 1350 TVVIGNDGAVD-PSVVGQTFKDGAGLRFYNLKGQLVGTSGWYETANHDWVYIQ----- 1404
Db 1595 SYANRNDGLFLNAPVQVKGQYLAGNSN-----QYKGOQVTTIAGVANVSGKDWLSISNGT 1649
Qy 1405 -----SGKALTEBOT-----INGOHLY--PKEDGHQVKG-----OLVTGT 1437
Db 1650 QYWIDSQLANTFNTHDMNQKVFVNNTSNLGLFLNAPYRQPGYKLAGLAKNNYNTQVTVS 1709
Qy 1438 DGKRYRYDANGSDQAFNKS-VTVNGKTYVFGNDGTAGTGNPKGTGFKGSDIR----- 1490
Db 1710 Q-----QYFD-----DQGTWVSQVVLGGQVVDNHALAQVSDTDQQLVYNSNGRNDGLFL 1762
Qy 1491 ---FYSMEGQVLTSGWYENAGQWLYV-KNGKVLTLGLQ-----TVGSRVYFDENGI 1539
Db 1763 NPYRGQGSQLI---GWTADYNGHVQVTKGQDAYGAQWFLIILNQQVWVDSRAL 1816

RESULT 11
US-09-604-957-3
; Sequence 3, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1278
; TYPE: PRT
; ORGANISM: Lactobacillus reuteri
US-09-604-957-3

Query Match 27.4%; Score 2284; DB 2; Length 1278;
Best Local Similarity 46.0%; Pred. No. 7.2e-152;
Matches 497; Conservative 167; Mismatches 338; Indels 78; Gaps 29;

Qy 116 VPTENEN-----QGFTDEMLAEAKNVATAESDSIP-----SDLAKMSNVKQVDGKYY 163
Db 185 VTKNENVLVHRFSNDVKTGEGNYVDWFSELMPVKDSFOKNGPLKQFGLQTINGQYY 244
Qy 164 YD-QDGNVKNFAVSGDKIYYFD-ETGAYKDTSKVDADKSSAVSQNATIFAAANNRAYS 221
Db 245 IDTTGQPRKNFLQSGNWIYFSDTGV--GTNALELQAKGTVSSNEQ-YRNGNAYS 301
Qy 222 TSAKNFEAVDNYLTADSWYRPSKILDKGKTWTSKDDPRFLMAMWPDTEKRYNYVM 281
Db 302 YDKSIEVNGYLTAADTYRKPQLKDGTTWTDSTKETDMEPILMVMWPNLTQAYLYNM 361
Qy 282 ----NKVVGIDKTYTAETSQADLTAAELVQARIQKITSSENTKWLREISAIFVKTQPP 337
Db 362 KQHGNLLPSALPFNADADPAELNHYSEIVQOQIEKRISSETGTMTDLRLTLMHDFVTNNPM 421
Qy 338 WNGESEKPYDD--HLQNGALLFNQDTLTPDTCOSYRLLNRTPTNQTGSLDSRFTYNNPD 395
Db 422 WNKDSNVNPSGTFQGGFLKYEN-SDLTPYANSYRLGLRMPINIKQ-----TYR--- 472
Qy 396 PLGGYDFLLANDVDSNPVQAEQLNWLHLLNFGSIYANDADANFDSIRVDVAVNDAD 455
Db 473 ---QGEFLANDIDNSNPVQAEQLNWLHLLNFGITANNQDANFDSVRVDAPNDAD 529
Qy 456 LQOISSDYKAAVIGDKNNKANNHYSEIVAEWSDNTPYLHDGDNLMNMNDFRLSLMLW 515
Db 530 LAMIAQDYFNAAYGMD-SOAVSNKHINILEDNMHADPEYFNKIGNPQLTMDDTIKNSLNH 588

Qy 516 SLAKPLDKRSLGNPLIHNSLVDRVDDREVBETVPSYSFARAHDSVQDIIRDIILKABINP 575
Db 589 GLSDATN-RWGLDAIVHQSLADRENNSTENVVIENYSFVRADNNNSQDIQNAIR-DVTG 646
Qy 576 NSFYSFTQBEIEQAFKIYNEDLKKTKYTHYNVPLSYTLTLTNKGSIPRVYTGDMFTD 635
Db 647 KDY-HTTFEDEQIGIDAYIQDQNSVTVKYNLYNIPASYAILLTNKTDTIPRVYIGDLVTD 705
Qy 636 DGYMANKTVMYDAIESILLKARKMYVSGGQAMQYQI--GNGEILTSVRYGKALKQSKDG 694
Db 706 GGOYMEHQTRYDYTLTLNLSRVKYVAGGOSMOTMSVGGNNNLTLSVRYGKAMTATDTG 765
Qy 695 DATRTTSGVGMGNQPNFSL--DGKVVALNMGAHANOEYRALMNVSTKCGVATYATDAD 752
Db 766 TDETRTQIGVYVSVNTENLKGVDKVV-LHMGAAHKNQYRAAVLTITDGVINYTSQGG 824
Qy 753 ASKAGLVKRTDENGLYFLNDDL-----KGVANPQVSGFLQVWVPVGAADQD 800
Db 825 AP-----VAMTDENGDLVLSHNLVNGKEADTAVQGYANPDSVGYLAVWVPVGAASNDQD 880
Qy 801 IRVAASDTASTDGKSLHODAMDSRVMEGFSNFSQSFATKEEETNVVIANNVDFVSWG 860
Db 881 ARTAPSTEKSGNSAYRTNAAFDSNVIFEAFSNFVYPTTKESERANVRIAQNADFFASLG 940
Qy 861 ITDFEMAPQVSVSTDGQFLDSVIONGYAFTRDYDLGMSKANKYGTADOLVKAIKALHAKG 920
Db 941 FTSFEMAPQVNSKDRFTLSDTDNGYAFTRDYDLGMSKANKYGTADOLVKAIKALHAKG 1000
Qy 921 LKVMADWVPDQMYTFPKQEVYVTRTDKFGKTAGSQINHSLYVYVTDKSSGDDYQAKYGG 980
Db 1001 LQVMADWVPDQIYNLPGKEVATVTRVDDRGNVKWDALINNLYVNT-IGGGEVQKYYGG 1059
Qy 981 AFLDELKEKPELFTKKQISTQAIQIDPSVKIKWSAKYFNGSNILRGADYVLSQDVSNK 1040
Db 1060 AFLDKLQKLYEFTTKQVSTGVAIDPSQKITMSAKYFNGSNILRGADYVLSQDVSNK 1118
Qy 1041 YFNVASDT-LFLPSSLLGKVVESGIRY---DGKGYLYNSSATGQDVKASFITEA-GNLY 1094
Db 1119 YNLTGTTTKQFLPIQLTGKQKQNEGFVKNGDNGYFYD--LAGNMVKNFTFIEDSVGNWY 1176
Qy 1095 YFGKGYVMVTGAQTIN-----GANYFFLENGTALR-NTIYTDAGNNSHYVANDGKRYEN 1147
Db 1177 FPDQDGKGVENKHFVDVDSYGEKGTFFLKNKGVSRGGLVQTD--NGTYIFDNYGKAVRN 1234

RESULT 12
US-09-995-749A-2
; Sequence 2, Application US/09995749A
; Patent No. 6867026
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: BO43388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1781
; TYPE: PRT
; ORGANISM: Lactobacillus reuteri
US-09-995-749A-2

Query Match 27.4%; Score 2284; DB 2; Length 1781;
Best Local Similarity 46.0%; Pred. No. 1.2e-151;

Matches	497;	Conservative	167;	Mismatches	338;	Indels	78;	Gaps	299
Qy	116	VPTENEN----	QGFTDEMLAEAKNVATAESDSIP-----	SDLAKMSNVKQVDGKYYY	163				
Db	688	VTVKENVQLVHRFSNDVKTEGEGYVDFWSELMPVKDSFQKGNGLPKQFGLQTINGQYY	747						
Qy	164	YD-ODGNVKNFVAVSGDKIYYFD-ETGAYKOTSVDADKSSSASVQATIFPAANRRAYS	221						
Db	748	IDPTTGQPRKNFLILQSGNNWIYFSDSDTGV--GTNALELQFAKGTVSNEQ-YRNGNAAYS	804						
Qy	222	TSAKNFEADVNYLTADSWYRPSKSLDKGKTWTSRGKDDFPFLMAWMPDETETKXNVNYM	281						
Db	805	YDDKSIENVNGYLTDYTPKQILDKGTITWTSKETDMRPILMVMWPNLTLLQAYLYNM	864						
Qy	282	---NKVVGIDKTVTAETSOADLTAAAEVQARIEQKITSENNWKMLREASIAFVKTPQ	337						
Db	865	KQHCNLLPSALPFPNADADPAELNHYSEIYQOHIKRISETGNTDMLRTLMLHDFVTNPM	924						
Qy	338	WGESEKPYDD--HLQNGALLFNOQTDLTPTDTSYRLLNRTPTNQTGSLDSRFTYNPND	395						
Db	925	WNKDSNVNFGSQFQGGFLKYEN-SDLTPVANSDYRLGRMPINIKDQ-----TYR--	975						
Qy	396	PLGGYDFELLANDVDSNPVQAQOLNWLHYLLNFGSIYANDADANPDSIRVDAVDNVAD	455						
Db	976	---GQEFELLANDINSNPVQAQOLNWLHYLLNFGTITANDQANFDSVRVADPNIDAD	1032						
Qy	456	LLQTSDDYLKAAYGIDKNNKANNHVSIVBAWSDNDTPYLHDDGDLMMNMNDFRLSMILW	515						
Db	1033	LMNIADQYFNAAYGMD-SDAVSNKHINILEDWHADPEYFNKIGNPOLTWDDTIKUSLNH	1091						
Qy	516	SLAKPLDKRGLNPLIHNSLVDRVEDDREYVETVPYSIFARAHDSVEQDIIRDIKAENP	575						
Db	1092	GLSDATN-RWGLDAIVHQSADRNNSTENVVFNYSFVARHDNNSDQIQNAIR-DVTG	1149						
Qy	576	NSFGYSTQBEIEQAFKIYNEDLKKTDKKYTHNVPLSYLLTNKGSIRPRVYVGMFTD	635						
Db	1150	KDY-HTFTFDEBEQKIDAYIQDQNSTVYKTKYLNIPASAYAILLTNKTOTIERVYVGDLYTD	1208						
Qy	636	DGYVMANKTVNYDAIESLLKARKMYVSGSGAQMNYQI-NGGEILTSVRYCKGALKQSDKG	694						
Db	1209	GGQYMEHQTRYDYLTLNLLKSRVKYVAGGSGMQTMSVGGNNILTSVRYCKGAWTADTG	1268						
Qy	695	DATRTTSGVGMGNQPNFSL--DGKVALNMGAAHANQRYALMWSTDKGVATYATDAD	752						
Db	1269	TDERTTQIGVGVVSTNTPLNKLGVNDKVV-LHMGAAHKNQYRAAVLTITTDGVINYTSQ	1327						
Qy	753	ASKAGLVKRTDENGYLFLNDDL-----KGVANPQVSGFLQVWVPVGAADDQD	800						
Db	1328	AP----VAMTDENGDLTSLSHNLVYNGKEADTAVQGYANPDVSGYLAVWVPVGAASDQ	1383						
Qy	801	IRVAASDTASTDGKSLHQDAAMDSRVMEFGSPQSFATKEEYTNVVVIANNVDKFFVSWG	860						
Db	1384	ARTAPSTKNGSNGSAYRTNAAPDSNVIFEPASNVTYPTTKESRANVRIAQNADFFASLG	1443						
Qy	861	ITDPEMAPQVYSTDQGLDSVIONGYAFTRDYDLGMSKANKYCTADQLVKAIKALHAKG	920						
Db	1444	FTSEMAPQYNSSKDRFTLSTIDNGYAFTRDYDLGNSFPNKYCTDDELRNALCALHAKG	1503						
Qy	921	LKYNADWVPQMYTFPKQEVVTVTRTDKFKPIAGSQIHNLSLYVYDTKSSGGDDYQAKYGG	980						
Db	1504	LQVMADWVPQIYNLPCKEVATVTRVDRDRGNVWKDAIINNLYVYNT-IGGGEYQKYYGG	1562						
Qy	981	AFIDELKEKYPELFTKKQISTGQAI DSPVKIKOWSAKYFNGSNLTILRGADYVLSDQVSNK	1040						
Db	1563	AFIDKLQKLYPELFTKKQVSTGVVAIDPSQKITEWSAKYFNGTINILHRGSGYVLKAD-GGQ	1621						
Qy	1041	YFNVASDT-LFLPSSILKGVVSEGRY----DGKGYLYNSSATGDQVKASFITEA-ONLY	1094						
Db	1622	YYNLGITTKQLPLQLTGKKQKQNEGPFVKNGDNGNYEYD-LAGNVKNTFIEDSVGNWY	1679						
Qy	1095	YFGKDGVMYVTAQTFIN-----GANYPFLBNGTALR-NTIYTDQAQNSHYVANDGKRYEN	1147						
Db	1680	FFDQDGKVMKNGKHFPVDVDSYGEKTYTFFLKNGVYSFRGLQVQTD--NCTYVYFDNYGKVMRN	1737						

```

RESULT 13
US-09-604-957-4
; Sequence 4, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEEB, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCES: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-604-957-4

Query Match          19.8%  Score 1653; DB 2; Length 545;
Best Local Similarity 57.6%; Pred. No. 5.7e-108;
Matches 315; Conservative 94; Mismatches 118; Indels 20; Gaps 5;

Qy  403  LLANDVNSNPVQAQLNLWHLNLFSGSIYANDADANFDSIRVDADVNDALQIISD 462
Db  1  LLANDINSNPVQAQLNLWHLNMYGSIANDPEANFDGVRVDADVNDALQLIASD 60
Qy  463  YLKAAYGIDKNNKNNANNHVSIVSAWSNDTPPYLHDDGDNLMMNDNFRLSMLWSLAKPLD 522
Db  61  YLKAHYGVDSKSEKNAIHLSEAWSNDNPQYNKDTKGAQLPIDNKLRLSLLYALTPELE 120
Qy  523  K-----RSGNLPLIHNSLVDREVDDREVTVPVSPYSFARAHDSVQDIIRDIKAEIN 574
Db  121  KDAENKNEIRSGLEPVIITNSLNRSAEGKNSERMANYPFIRAHDSVQTVIAKIIKAQIN 180
Qy  575  PNAFGYSFTOEIEQAFKINEDLKTKDKKYTHYNVPLSYTLTLTKNGSIPRVVYGDMT 634
Db  181  PKTDGLTFTLDELKQAFKINEDMRQAKKYKKTOSNPTAYALMLSNKDSITRLYGDMS 240
Qy  635  DGGYMANKTVMYDAIESLLKARKMYVSGQAMQ-NYQIONGE-----ILTSVRYGK 685
Db  241  DDGYMATKSPYDAIDTLLKARIKYAAGQDMKITVEGDGSHMDWDYTGVLTSVRYGT 300
Qy  686  GALKQSDKGDATRTTSGVGVVMGNQPNFSLD-GKVVALNNGAAHQAHOEYRALMVSTKGV 744
Db  301  GANEATDQGEARTQTGMVITSNFSLKNQNDKVIIVNNGAAHQAHOEYRPLLTUKDGL 360
Qy  745  ATYATDADASKAGLVKRTDENGLYLFLNDDLKGVANPQVSGFLQVWPVGAADDQDIRVA 804
Db  361  TSYTSDAAAK--SLYRKTNDKGLVFPDASDIQGYLNPQVSGYLAVWPVVGASDNQDVRVA 418
Qy  805  ASDTASTDGLSKLHQDAADRVRVMEFESNFQS FATKEEYETNVVIANNVKDFVSWGITDF 864
Db  419  ASKANATGVQVYESSALDSQLIYEGFSNFPQDFVTKDSYTNKKIAQNVLQFKSWGVTSF 478
Qy  865  EMAPQVYSSTGDGFLDSVIQNGYAFTRDYDLGMSKANKYGTADQLKVAIKALHAKGLKVM 924
Db  479  EMAPQVYSSEGGFLDSIIQNGYAFEDRYDLAMSKNNKYGSQQDMINAVKALHKSGLQVI 538
Qy  925  ADWVPDQ 931
Db  539  ADWVPDQ 545

```

RESULT 14
US-09-995-749A-10
; Sequence 10, Application US/09995749A

```
; Patent No. 6867026
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: B043388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Streptococcus mutans
; US-09-995-749A-10

Query Match      19.5%; Score 1631; DB 2; Length 545;
Best Local Similarity 57.5%; Pred. No. 2e-106;
Matches 315; Conservative 94; Mismatches 117; Indels 22; Gaps 7;

QY 403 LLANDVNSNPVVOAEQLNWLHLLNFGSIYANDADANFDSIRVDVNDVADLLQISSD 462
Db 1 LLANDIDNSNPVVOAEQLNWLHLLNFGSIYANDPEANFDGVRVDVNDVADLLQIAD 60
QY 463 YLKAAYGIDKNKNNHVSIVAEASNDNTPYLHDDGDNLMNMDKFRSLMWSLAKPLD 522
Db 61 YLKAHYGVDPKSEKNAHLSLEAWSNDPQYNKDTKGALPTIDNKLRLSLLYALTRPLE 120
QY 523 K-----RSGLNPLHNSLDREVDDREVETVPSYFARAHDSVQDIIRDIKAEIN 574
Db 121 KQASNKNEIRSGLEPVITSLNRSAGKNSERMANYIFIRAHDSVQTVIAKIIKQAIN 180
QY 575 PNSFGYSFTQEEIQAFKIYNEDLKKTKYTHYNVPLSYTLTLTNKSGIIPRVYGDMT 634
Db 181 PKTDGLTFLDELKQAFKIYNEDMRQAKKKYQTSNIPTAYALMLSNKDSITRLYGDMTS 240
QY 635 DGGYWMANTVNDAYIESLLKARKMYSGGQAMQ-NYQINGE-----ILTSVRYGK 685
Db 241 DGGYMATKSPYYDAIDTLLKARIKAYAGQDMKITVEGDKSHMDWDTGVLTSVRYGT 300
QY 686 GALKOSDKGATRTSGVGMGNQPNFSLD-GKVVALNMGAAHANOEVYRALMVSTKDG 744
Db 301 GANEATDQGEATKTQGMVITSNFSLKLNQNDKRVIVNMGAAHKNQYRPLLLTTKDLG 360
QY 745 ATYATDADAKAGLVKRTDENGILYFLNDDLKGV-ANPQVSGFLOVWVPVGAADQDIRV 803
Db 361 TSYTSDAAK--SLYRKTNDKGLVFDASDIQGLYLNQVSG-LAWVPVGSADNQDIRV 417
QY 804 AASDTASTDGSLHDOADMSRMFEGFNFSQFATKEEYTNVVIANNVDFVSGMGT 863
Db 418 AASNKANATGQVYESSSALDSQLIYEGFSNFQDFVTKDSDTYNTKKIAQNQLFKSGVTS 477
QY 864 FEMAPQVYSTDQGLDSVIQNGYAFTRDYDLGMSKANKYGTADQLVKAIKALHAKGLV 923
Db 478 FEMAPQVYSSDGSFLDSIIQNGYAFEDRYDLAMSNNKYNKYGSGQDMINAVKALHKS 537
QY 924 MADWVPDQ 931
Db 538 IADWVPDQ 545
```

```
RESULT 15
US-09-604-957-5
; Sequence 5, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
```

```
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: B043388
; CURRENT APPLICATION NUMBER: US/09/604,957
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 523
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
; US-09-604-957-5
```

```
Query Match      19.2%; Score 1600; DB 2; Length 523;
Best Local Similarity 58.3%; Pred. No. 2.9e-104;
Matches 311; Conservative 76; Mismatches 132; Indels 14; Gaps 6;

QY 403 LLANDVNSNPVVOAEQLNWLHLLNFGSIYANDADANFDSIRVDVNDVADLLQISSD 462
Db 1 LLANDVNSNPVVOAEQLNWLHLLNFGSIYANDADANFDRVDAVNDVADLLQIAD 60
QY 463 YLKAAYGIDKNKNNHVSIVAEASNDNTPYLHDDGDNLMNMDKFRSLMWSLAKPLD 522
Db 61 YKFLAYGVDPDNDATANHLSLEDNSHNDPLVYTDGNSQLTMDYVHTQLINSLTKSSD 120
QY 523 KSGLNPLHNSLDREVDDREVETVPSYFARAHDSVQDIIRDIKAEINP---NSFG 579
Db 121 IRGTWQRFVYDMVDRSNDSTENEAIPNYSFVRAHDSVQTVIAQIV-SDLYPDVENS- 178
QY 580 YSFTQEEIQAFKIYNEDLKKTKYTHYNVPLSYTLTLTNKSGIIPRVYGDMDTDDG 639
Db 179 -APTEQLAAAPKVYNDEKLDKKTQYNMASAYAMLLTNKDTVPRVYGDLYTDDGQY 237
QY 640 MANKTVNYDAIESLLKARKMYSGGQAMQNYQINGEILTSVRYGKALKQSDKGATTR 699
Db 238 MATKSPYYDAINTLLKARQYVAGQSM---SVDSNDVLTSVRYGKDATASDTGTSETR 294
QY 700 TSGVGMGNQPNFSL-DGKVVALNMGAAHANOEVYRALMVSTKOGVATYATDADASKAGL 758
Db 295 TEGIGVIVSNNAELQLEDGHTVTLHMGAAHKNQYRALLSTTADGLAYVYDTDENAP- 350
QY 759 VKRTDENGILYFLNDDLKGVANPQVSGFLOVWVPVGAADQDIRVVAASDTASTDGSL 818
Db 351 VAYTDANGDLIFTNBSIYGVQNPQVSGYLAVWVPVGAQQDDARTASDTTNTSDKVFS 410
QY 819 DAAMDSRVMEFGFSNFQSFATKEEYTNVVIANNVDFVSGMGTDPFEMAPQVYSSDQ 878
Db 411 NAAALDSQVIYEGFSNFQAFATDSSSEYTNVVIANNADQFKQNGVTSFQLAPQVRS 470
QY 879 LDSVTQNGYAFTRDYDLGMSKANKYGTADQLVKAIKALHAKGLKVMADWVPDQ 931
Db 471 LDSIIQNGYAFTRDYDLGVTGTPKYGTADQLRDAIKALHASGIQAIADWVPDQ 523
```

Search completed: February 11, 2006, 19:45:54
Job time : 42.8119 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
 Copyright (c) 1993 - 2006 Bioceleration Ltd.
 OM protein - protein search, using sw model
 Run on: February 11, 2006, 20:39:23 ; Search time 142.315 Seconds
 (without alignments)
 4668.162 Million cell updates/sec
 Title: US-10-797-821-37
 Perfect score: 8349
 Sequence: 1 MEKNVRFKMKVKRWVTLUS.....VYQYYFFGNDGARIYRGMN 1590
 Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5
 Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : Published Applications AA_Main:*

- 1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pep:*
- 3: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pep:*
- 4: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pep:*
- 5: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pep:*
- 6: /cgn2_6/prodata/1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8349	100.0	1590	4	US-10-383-930-37
2	8349	100.0	1590	5	US-10-797-821-37
3	4580	54.9	1475	4	US-09-740-274-2
4	4580	54.9	1475	4	US-10-383-930-34
5	4580	54.9	1475	5	US-10-797-821-34
6	4214.5	50.5	1375	3	US-09-740-274-4
7	4214.5	50.5	1375	4	US-10-383-930-35
8	4214.5	50.5	1375	5	US-10-797-821-35
9	3575	42.8	1430	3	US-09-740-274-6
10	3575	42.8	1430	4	US-10-383-930-36
11	3575	42.8	1430	5	US-10-797-821-36
12	3502.5	42.0	1554	4	US-10-383-930-38
13	3502.5	42.0	1554	5	US-10-797-821-38
14	3325.5	39.8	1518	4	US-10-383-930-40
15	3325.5	39.8	1518	5	US-10-797-821-40
16	3031	36.3	1365	4	US-10-383-930-39
17	3031	36.3	1365	5	US-10-797-821-39
18	2975	35.6	1497	5	US-10-484-218-18
19	2808	33.6	1595	5	US-10-484-218-20
20	2472	29.6	1006	5	US-10-484-218-22
21	2454	29.4	2057	4	US-10-417-280A-2
22	2292.5	27.5	1777	5	US-10-484-218-12
23	2284	27.4	1781	3	US-09-995-749A-2
24	2220.5	26.6	1771	5	US-10-484-218-14
25	1631	19.5	545	3	US-09-995-749A-10
26	1595.5	19.1	522	3	US-09-995-749A-11
27	1401	16.8	535	3	US-09-995-749A-13

28 1368.5 16.4 787 5 US-10-484-218-16 Sequence 16, Appl
 29 1260.5 15.1 584 3 US-09-995-749A-12 Sequence 12, Appl
 30 1206 14.4 525 5 US-10-484-218-23 Sequence 23, Appl
 31 675 8.1 223 5 US-10-484-218-6 Sequence 6, Appl
 32 669.5 8.0 224 5 US-10-484-218-4 Sequence 4, Appl
 33 653 7.8 223 5 US-10-484-218-10 Sequence 10, Appl
 34 553 6.6 221 5 US-10-484-218-2 Sequence 2, Appl
 35 543 6.5 221 5 US-10-484-218-8 Sequence 8, Appl
 36 514 6.2 2710 4 US-10-011-366-6 Sequence 6, Appl
 37 514 6.2 2710 4 US-10-354-774-6 Sequence 6, Appl
 38 514 6.2 2710 4 US-10-271-012-6 Sequence 6, Appl
 39 514 6.2 2710 4 US-10-729-122-6 Sequence 6, Appl
 40 514 6.2 2710 4 US-10-729-039-6 Sequence 6, Appl
 41 514 6.2 2710 5 US-10-729-527-6 Sequence 6, Appl
 42 514 6.2 2710 5 US-10-727-898-6 Sequence 6, Appl
 43 514 6.2 2710 5 US-10-728-696-6 Sequence 6, Appl
 44 514 6.2 2710 6 US-11-001-241-6 Sequence 6, Appl
 45 459 5.5 866 4 US-10-222-038-2 Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-10-383-930-37
 ; Sequence 37, Application US/10383930
 ; Publication No. US20040127400A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Smith, Daniel J
 ; APPLICANT: Taubman, Martin A
 ; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
 ; FILE REFERENCE: 25669-018
 ; CURRENT APPLICATION NUMBER: US/10/383,930
 ; CURRENT FILING DATE: 2003-03-07
 ; PRIOR APPLICATION NUMBER: 60/402,483
 ; PRIOR FILING DATE: 2002-08-08
 ; PRIOR APPLICATION NUMBER: 60/363,209
 ; PRIOR FILING DATE: 2002-03-07
 ; NUMBER OF SEQ ID NOS: 41
 ; SOFTWARE: Patent version 3.2
 ; SEQ ID NO 37
 ; LENGTH: 1590
 ; TYPE: PRT
 ; ORGANISM: Streptococcus sobrinus
 US-10-383-930-37

Query Match 100.0%; Score 8349; DB 4; Length 1590;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1590; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEKNVRFKMKVKRWVTLUSASATMLASALGASVASADTDATDQASAEQTTGTTASTDTAAQTNTTANAEKVPYPTEN 60
 DB 1 MEKNVRFKMKVKRWVTLUSASATMLASALGASVASADTDATDQASAEQTTGTTASTDTAAQTNTTANAEKVPYPTEN 60

QY 61 QATDQTSIAATATSEQASSTDAATDQASAEQTTGTTASTDTAAQTNTTANAEKVPYPTEN 120
 DB 61 QATDQTSIAATATSEQASSTDAATDQASAEQTTGTTASTDTAAQTNTTANAEKVPYPTEN 120

QY 121 ENQGFTEMLAKNKNVATAESDIPSDLAKNKNVQVGGYKYVYDQDGNVKNFVSVGD 180
 DB 121 ENQGFTEMLAKNKNVATAESDIPSDLAKNKNVQVGGYKYVYDQDGNVKNFVSVGD 180

QY 181 KIYIFDETGA YKDTSKVDADKSSAVSONATIPANNRAYSTSAKNFEADVNTLTADSWY 240
 DB 181 KIYIFDETGA YKDTSKVDADKSSAVSONATIPANNRAYSTSAKNFEADVNTLTADSWY 240

QY 241 RPKSILKDGKTWTEGSKDDFRPLLMWPDTEKRYVNMKNKVGIDKTYTAETSQADL 300
 DB 241 RPKSILKDGKTWTEGSKDDFRPLLMWPDTEKRYVNMKNKVGIDKTYTAETSQADL 300

QY 301 TAAAEVLQARIPOKITSENNTKWLREAIISAFVKTPQWNGESEKPYDDHLQNGALLFDNQ 360
 DB 301 TAAAEVLQARIPOKITSENNTKWLREAIISAFVKTPQWNGESEKPYDDHLQNGALLFDNQ 360

QY 361 TDLTPDQSNRYLLNRTPTNQTGSLDSRFTYFNDPLGGYDFLLANDVDNSNPVVQAEQL 420
DB TDLTPDQSNRYLLNRTPTNQTGSLDSRFTYFNDPLGGYDFLLANDVDNSNPVVQAEQL 420
QY 421 NMLHYLLNFGSIYANDADANFDSIRVDADVNDADLLQISSDYLLKAAAYGIDKNNKNNHH 480
DB 421 NMLHYLLNFGSIYANDADANFDSIRVDADVNDADLLQISSDYLLKAAAYGIDKNNKNNHH 480
QY 481 VSIWEAWSNDPTPLHDDGDNLMNDKFLSLMLSLAKPLDKRSGNLPLIHNSLDVREV 540
DB 481 VSIWEAWSNDPTPLHDDGDNLMNDKFLSLMLSLAKPLDKRSGNLPLIHNSLDVREV 540
QY 541 DREVEVTPSYSPARAHSEVDQIIRDIKAEINPNSFGYSPTQEBIEQAFKIYNEDLKK 600
DB 541 DREVEVTPSYSPARAHSEVDQIIRDIKAEINPNSFGYSPTQEBIEQAFKIYNEDLKK 600
QY 601 TDKKYTHYNVPLSYTLILLTNKSGIPRVYVYGDMPFTDDGQYMANKTVNYDAIESLLKARMKY 660
DB 601 TDKKYTHYNVPLSYTLILLTNKSGIPRVYVYGDMPFTDDGQYMANKTVNYDAIESLLKARMKY 660
QY 661 VSGGQAMQYQIENGEBILTSVRKYGKALKQSDKGDATTRTSGVGVVMGNQPNFSLDGKVV 720
DB 661 VSGGQAMQYQIENGEBILTSVRKYGKALKQSDKGDATTRTSGVGVVMGNQPNFSLDGKVV 720
QY 721 ALNMGAHAHQEYRALMVSTKOGVATYATDADASKAGLVKRTDENGLYFLNDDLLKGVAN 780
DB 721 ALNMGAHAHQEYRALMVSTKOGVATYATDADASKAGLVKRTDENGLYFLNDDLLKGVAN 780
QY 781 PQVSGFLQVWPVGAADDODIRVAASDTASTDGKSLHODAAMDSTRVMEGFSNFQSFATK 840
DB 781 PQVSGFLQVWPVGAADDODIRVAASDTASTDGKSLHODAAMDSTRVMEGFSNFQSFATK 840
QY 841 EBEYTNVVIANNVDKPVSGNIGITDFEMAPQVVSSTDGQFLDSVTQNGYAFTRDYDLGMSKA 900
DB 841 EBEYTNVVIANNVDKPVSGNIGITDFEMAPQVVSSTDGQFLDSVTQNGYAFTRDYDLGMSKA 900
QY 901 NKYGTADQLVKAIFKALHAGLKVMAADWPDQMTFFPKQEVVTVTRTDKFKPIAGSQIINH 960
DB 901 NKYGTADQLVKAIFKALHAGLKVMAADWPDQMTFFPKQEVVTVTRTDKFKPIAGSQIINH 960
QY 961 SLVYVTDKSGDDYQAKYGAFDELKEKYPELFTKKQISTGQAIIDPSVKIKQWSAKYFN 1020
DB 961 SLVYVTDKSGDDYQAKYGAFDELKEKYPELFTKKQISTGQAIIDPSVKIKQWSAKYFN 1020
QY 1021 GSNILGRGADYVLSDOVSNKYFNVASDTLFLPSSLILGKVVEGIRYDGGKIYVNSSATGD 1080
DB 1021 GSNILGRGADYVLSDOVSNKYFNVASDTLFLPSSLILGKVVEGIRYDGGKIYVNSSATGD 1080
QY 1081 QVKASPFITEAGNLYYFGKDGVMYVTAQTINGANYFFLENGTALRNTIYTDACQNSHYAN 1140
DB 1081 QVKASPFITEAGNLYYFGKDGVMYVTAQTINGANYFFLENGTALRNTIYTDACQNSHYAN 1140
QY 1141 DGKRYENGYQOFGNDHRYPKDGNMAGLITVDGNVQYFDKGVQAKDKIIVTRDGVKRYF 1200
DB 1141 DGKRYENGYQOFGNDHRYPKDGNMAGLITVDGNVQYFDKGVQAKDKIIVTRDGVKRYF 1200
QY 1201 DOHNGNAAWTNFTADKTGHWYLGKDGVAVTGAQTVGKOKLYPEANGQOVKGDFTVSDG 1260
DB 1201 DOHNGNAAWTNFTADKTGHWYLGKDGVAVTGAQTVGKOKLYPEANGQOVKGDFTVSDG 1260
QY 1261 KLYFYDVDSGDMWTDFTIEDKAGNWFYLGKDGAAVTAQTIRGQKLYFKANGQOVKGDIV 1320
DB 1261 KLYFYDVDSGDMWTDFTIEDKAGNWFYLGKDGAAVTAQTIRGQKLYFKANGQOVKGDIV 1320
QY 1321 KGTGDKIRYDAGSGQVFNKTKAADGKTYVTCNDGVAVDPSVVKQTFKDGASGALRFY 1380
DB 1321 KGTGDKIRYDAGSGQVFNKTKAADGKTYVTCNDGVAVDPSVVKQTFKDGASGALRFY 1380
QY 1381 NLKQGLVTSQGWYETANHDWVYIQSGKALTGEOTINGOHLYFKEDGHQVKGQLVGTGDGK 1440
DB 1381 NLKQGLVTSQGWYETANHDWVYIQSGKALTGEOTINGOHLYFKEDGHQVKGQLVGTGDGK 1440

QY 1441 VRYDANSQDQAFNKSVTTVNGKTYFFGNDGTAQTAGNPKGQTFKQSGDIRFYSMEQOLVT 1500
DB 1441 VRYDANSQDQAFNKSVTTVNGKTYFFGNDGTAQTAGNPKGQTFKQSGDIRFYSMEQOLVT 1500
QY 1501 GSGWENAGQOWLYYKNGKVLTLGLOTVGSORVYFBEINGIQAAGKAVRTSDGKIRYFDENS 1560
DB 1501 GSGWENAGQOWLYYKNGKVLTLGLOTVGSORVYFBEINGIQAAGKAVRTSDGKIRYFDENS 1560
QY 1561 GSMITNQMKFVYGYQYFFGNDGARIYRGNW 1590
DB 1561 GSMITNQMKFVYGYQYFFGNDGARIYRGNW 1590

RESULT 2
US-10-797-821-37
; Sequence 37, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 37
; LENGTH: 1590
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-797-821-37

Query Match 100.0%; Score 8349; DB 5; Length 1590;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1590; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MEKNVRFKHKYKKRWVTLVSASATMLASALGASVASADTDTASDDSNQAVVTGDDTTNN 60
DB 1 MEKNVRFKHKYKKRWVTLVSASATMLASALGASVASADTDTASDDSNQAVVTGDDTTNN 60
QY 61 QATDQTSIAATATSBQASASTDAATDQASAAEQTQGTASTDTAAQTNNANEAKWPTEN 120
DB 61 QATDQTSIAATATSBQASASTDAATDQASAAEQTQGTASTDTAAQTNNANEAKWPTEN 120
QY 121 ENQGFTEMLAEAKNVATAESDIPSDLAKMSNVKQVKGKYYYDQDGNVKNFVSVGD 180
DB 121 ENQGFTEMLAEAKNVATAESDIPSDLAKMSNVKQVKGKYYYDQDGNVKNFVSVGD 180
QY 181 KIYYFDETGAAYKOTSKVDADKSSSAVSQNAATIPANNRAYSTSAKNFEADVNLTDADSWY 240
DB 181 KIYYFDETGAAYKOTSKVDADKSSSAVSQNAATIPANNRAYSTSAKNFEADVNLTDADSWY 240
QY 241 RPKSILKQKWTESKDDPRLLMAWPDUTTKRNYVMNMKVVGIDKTYTAETSOADL 300
DB 241 RPKSILKQKWTESKDDPRLLMAWPDUTTKRNYVMNMKVVGIDKTYTAETSOADL 300
QY 301 TAAAEVLQARIQKITSSENTNKTWLRSAISAFVKTPQWNGESEKPYDDHLQNGALLFDNQ 360
DB 301 TAAAEVLQARIQKITSSENTNKTWLRSAISAFVKTPQWNGESEKPYDDHLQNGALLFDNQ 360
QY 361 TDLTPDQSNRYLLNRTPTNQTGSLDSRFTYFNDPLGGYDFLLANDVDNSNPVVQAEQL 420

Db 361 TDLTPDTQSNRYLLNTPNTQTSLSRSRTYFNPNDFLGGYDFLLANDVNSNPNVQAEQL 420
QY 421 NMLHYLLNFGSIYANDADANFDSIRVDADVNDVADLLOJSSDYLLKAAYGIDKNNKANHH 480
Db 421 NMLHYLLNFGSIYANDADANFDSIRVDADVNDVADLLOJSSDYLLKAAYGIDKNNKANHH 480
QY 481 VSIWEAWSNDTPYLHDDGDNLMNDKFLRLSLAKPLDKRSGLNPLIHNLSVDREV 540
Db 481 VSIWEAWSNDTPYLHDDGDNLMNDKFLRLSLAKPLDKRSGLNPLIHNLSVDREV 540
QY 541 DREVEVTPSYSPARAHSEVDIIRDIKAEINPNPSFGYFTQEBIEQAFKIYNEDLKK 600
Db 541 DREVEVTPSYSPARAHSEVDIIRDIKAEINPNPSFGYFTQEBIEQAFKIYNEDLKK 600
QY 601 TDKKYTHYNVPLSYTLNLLTNKSGIPRVYVYGDFTDGGQYMANNTVNYDAIESLLKARMKY 660
Db 601 TDKKYTHYNVPLSYTLNLLTNKSGIPRVYVYGDFTDGGQYMANNTVNYDAIESLLKARMKY 660
QY 661 VSGQAMQNYQIENGEBILTSVRYGKALKQSDKGDATTSTSGVGVVMGNQPNFSLDGKV 720
Db 661 VSGQAMQNYQIENGEBILTSVRYGKALKQSDKGDATTSTSGVGVVMGNQPNFSLDGKV 720
QY 721 ALNMGAHANOEYRALMWSTKDGVAITYATDADASKAGLVKRTDENGILYFLNDDLKGVAN 780
Db 721 ALNMGAHANOEYRALMWSTKDGVAITYATDADASKAGLVKRTDENGILYFLNDDLKGVAN 780
QY 781 PQVSGFQWVPVGAADDQDQIRVAASDTASTGKSLHQAAMDSRWPFEGFNFQSFATK 840
Db 781 PQVSGFQWVPVGAADDQDQIRVAASDTASTGKSLHQAAMDSRWPFEGFNFQSFATK 840
QY 841 EBEYTNVVIANNVDFVSWGITDFEMAPQYVSVSTDGFLDSVIQNGYAFTRDYDLGMSKA 900
Db 841 EBEYTNVVIANNVDFVSWGITDFEMAPQYVSVSTDGFLDSVIQNGYAFTRDYDLGMSKA 900
QY 901 NKYGTADQLVKAIKALHAKGLKVMADWPDQYTPPKQEVVTVTRTDKFGKPIAGSQINH 960
Db 901 NKYGTADQLVKAIKALHAKGLKVMADWPDQYTPPKQEVVTVTRTDKFGKPIAGSQINH 960
QY 961 SLVYVTDTKSSGGDYQAKYGGAFLDELKEYPELFTKQIISTGOAIDPSVKIKQWSAKYFN 1020
Db 961 SLVYVTDTKSSGGDYQAKYGGAFLDELKEYPELFTKQIISTGOAIDPSVKIKQWSAKYFN 1020
QY 1021 GSNILGRGADYVLSQVSNKYFNVASDTLFLPSSLGKVVESGIRYDGKGIYNSSATGD 1080
Db 1021 GSNILGRGADYVLSQVSNKYFNVASDTLFLPSSLGKVVESGIRYDGKGIYNSSATGD 1080
QY 1081 QVKASFITEAGNLYYFGKDGWMTGAQTINGANYFFLENGTALRNTIYTDAGNSHYAN 1140
Db 1081 QVKASFITEAGNLYYFGKDGWMTGAQTINGANYFFLENGTALRNTIYTDAGNSHYAN 1140
QY 1141 DGKRYENGYYQFQGNDRYFPKGNMAVGLTTVDGNVQYFDKGVQAKDKLIVTRDQKRVYF 1200
Db 1141 DGKRYENGYYQFQGNDRYFPKGNMAVGLTTVDGNVQYFDKGVQAKDKLIVTRDQKRVYF 1200
QY 1201 DOHNGNAATNTFIADKTHWYLYLGDGVAVTGAQTVGKQKLYFEANGQVKGDFVTSDBG 1260
Db 1201 DOHNGNAATNTFIADKTHWYLYLGDGVAVTGAQTVGKQKLYFEANGQVKGDFVTSDBG 1260
QY 1261 KLYFYDVDSGDMWTDFTIEDKAGNFWYLGKGAATVGAQIRGQKLYFKANGQQVKGDIV 1320
Db 1261 KLYFYDVDSGDMWTDFTIEDKAGNFWYLGKGAATVGAQIRGQKLYFKANGQQVKGDIV 1320
QY 1321 KGTGDKIRYDPAKSGEOPNKTVAADGKTYVIIGNDVAVDPSVVKGTQFKDASGALRPY 1380
Db 1321 KGTGDKIRYDPAKSGEOPNKTVAADGKTYVIIGNDVAVDPSVVKGTQFKDASGALRPY 1380
QY 1381 NLKQGLVTGSGHYETANHDWYIQSGKALTGBOTTINGHLYFKEDCHQVKGQVLTGTGK 1440
Db 1381 NLKQGLVTGSGHYETANHDWYIQSGKALTGBOTTINGHLYFKEDCHQVKGQVLTGTGK 1440
QY 1441 VRYDANSQDQAFNKSVTYNGKTYFFGNDGTAQTAGNPKGQTFKQSGDIRFYSMEGQLVT 1500

Db 1441 VRYDANSQDQAFNKSVTYNGKTYFFGNDGTAQTAGNPKGQTFKQSGDIRFYSMEGQLVT 1500
QY 1501 GSGWYENAGQWLYVXNGKVLTLGLQTVGSORVYFDENGIOAKGKAVRTSDGKIRYFDENS 1560
Db 1501 GSGWYENAGQWLYVXNGKVLTLGLQTVGSORVYFDENGIOAKGKAVRTSDGKIRYFDENS 1560
QY 1561 GSNITNMQKFVYQYVYFNGDGIARIYRGWN 1590
Db 1561 GSNITNMQKFVYQYVYFNGDGIARIYRGWN 1590
RESULT 3
US-09-740-274-2
; Sequence 2, Application US/09740274
; Patent No. US20020031826A1
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-740-274-2
Query Match 54.9%; Score 4580; DB 3; Length 1475;
Best Local Similarity 56.4%; Pred. No. 3.9e-242;
Matches 896; Conservative 212; Mismatches 338; Indels 144; Gaps 18;
QY 1 MEKNVRFKXHKVKKRWLTLSVASATMLASALGASVASADDTA-----SDSNQAVVTGQ 56
Db 1 MDKVRVYKLRKVKRWVTVSVASAVMTLTTLTSLGGLVKADSNESKQISNDSNTSVVTANE 60
QY 57 TTN--NQATDQTSIAATSEOSASTDAATDOASAAEQTGTTASTDTTAAQTITNANEAK 114
Db 61 ESNVTEATSEKQEAASSQTNHTVTVTSSSSTS VVNPKE-----VVSNPYTVGETASNGEKL- 115
QY 115 WYPTENENQGFDEMLABA-----KNVATASDSIPSD-----LAKMSNVKQVDGKY 161
Db 116 -----NQTTVDKTSSEAAANNISKQTEADTVIDSNAAANLQILEKLPNVEIDGKY 169
QY 162 YYYODGNVKNFVSVGDKIYYFDETGAUKDTSKVDADKSSSVASQNAITFAANNRAYS 221
Db 170 YYYDNNKVRNFTLIADGKILHFDETGAUTDTSIDTVNK--DIVTTRSNLYKKNQVYD 227
QY 222 TSAKNFEAVDNYLTADSNWRPKSILKDGKTWTSESKDDPRPLLMAWPDTETKRYNTVM 281
Db 228 RSAQSFEHVDHYLTAESWYRPKYILKDGKWTQSTEKDFRPLMTWMPDQETQRYNTVM 287
QY 282 NKVVGIDKTYTAETSQADLTAAAEVLQVARIKQITSENNTKWLREAIASFVKTOPOWNGE 341
Db 288 NALQINGINTYDDTSNQLQNTAAATIAQIEAKITILKNTDMLRQITIAFVKITQSANNSD 347
QY 342 SEKPYDDHLQNGALLFDNQDTLTPDTQSNRYLLNTPNTQTSLSRSRTYFNPNDFLGGYD 401

Db 348 SEKPFDDHLQNGAVLYDNEGKLTTPYANSNYRILNRTPTNQTGKDPRT--ADNTIGGYE 405
Qy 402 FLLANDVDNSNPVQAEQLNWLHLLNFSGSIYANDADANFDSIRVDADVNDADLLQISS 461
Db 406 FLLANDVDNSNPVQAEQLNWLHLLNFSGSIYANDADANFDSIRVDADVNDADLLQIAG 465
Qy 462 DYLLKAAAGIDKKNKANNHVSIVEAWSDDTPLYLHDDGDNLMNMONKFLSLMWSLAKPL 521
Db 466 DYLLKAAAGIHKNDKAANDHLSILEAWSDDTPLYLHDDGDNLMNMONKFLSLFLSLAKPL 525
Qy 522 DKRSGNLPLHNSLVDREVDREVEVTPSYSFARADHSEVDQIIRDIKAEINPNPSFGYS 581
Db 526 NORSGNPLITNSLVNRDNDNAETAAPVPSYFIRADHSEVDQIIRDIKAEINPNVUGYS 585
Qy 582 FTQEEIEQAFKIYNEDLKTKKYYTHYNVPLSYTLTLLTNKSGIPRYVYGDMTDDQYMA 641
Db 586 FTMEELKKAFAEYVNDLLATEKKYTHYNVPLSYTLTLLTNKSGIPRYVYGDMTDDQYMA 645
Qy 642 NKTVNYDAIESLLKARKVYSGGOAMONYQINGEILTSVRYGKALKOSDKGDATRTS 701
Db 646 HKTINYEALETLLKARIKIVSGGOAMRNOQVGNSEIITSVRYGKALKATDGDRTTTS 705
Qy 702 GGVVYMGNOPIFSLDGK--VVALNMGAAHANOEVRLMVSTKDGCVATYATDADASKAGLVK 760
Db 706 GVAVIEGNPURLKASDRVVNMGAHKNQAYRPLLLTDDNGIKAYHSDQEA--AGLYR 763
Qy 761 RTDENGILYFLNDDLKGVANPOVSGFLQVVPVGAADDQDIRVAASDSTATDCK--SLHQ 818
Db 764 YTNDRGELIFTAADIKGVANPOVSGVLGVVPVGA--LTKMFALLRLARPHQOMASVHQ 820
Qy 819 DAMDSRVMPFEGSFNFQSFATKEEYTNVIANNVKFKVSGWITDPEMAPQVVSSTDGOF 878
Db 821 NAALDSRVMPFEGSFNFQSFATKEEYTNVIANNVKFKVSGWITDPEMAPQVVSSTDGOF 880
Qy 879 LQSVIONGVAFTDRYDLGKSKANKYGTADOLVKAIKALHAKGLKVMADVPQMYTFPKQ 938
Db 881 LQSVIONGVAFTDRYDLGKSKANKYGTADOLVKAIKALHAKGLKVMADVPQMYTFPKQ 940
Qy 939 EYVTVTRTDKFGKPIAGSQINHSLYVTDTKSGDDYQAKYGAFLDELKEKPELFTFKQ 998
Db 941 EYVTVTRTDKFGKPIAGSQINHSLYVTDTKSGDDYQAKYGAFLDELKEKPELFTFKQ 1000
Qy 999 ISTGQADISVVKIKQWAKYFNGSNILGRADYVLSQVSNKYFVA--SDTLFLPSSL 1056
Db 1001 ISTGVPMDPSVKIKQWAKYFNGSNILGRADYVLSQVSNKYFVA--SDTLFLPSSL 1060
Qy 1057 GKVBESGIRYDGKGYLYNSSATGDQVKASFIIEAGNLYYFGKGYVMTGAQTINGANYEF 1116
Db 1061 NQDSQVGFSDGKGYVYIS--SGYQAKNTFISEGDKWYFDNNGYVMTGAQINGVNYF 1119
Qy 1117 LENGTLARNITYTDAQNSHYANDGKRYENGQOF--GNDWRYFKDGNMAVGLTTVDGNV 1175
Db 1120 LSNGLQLRAILLKNEGDYVAYGNDGRRYENGYYQPMGVRHFNNGEMSVGLTVIDGQV 1179
Qy 1176 QYFQDKGVQAKKIIIVTRDGKRYFPOHNGNAATNFIADKTHWYLYLKGDKGVAVTGAQT 1235
Db 1180 QYFDEMGYQAKGFVTTADKIRYFDKQSGNMYRNFIENEGKMWLYLGEDGAATVGSQT 1239
Qy 1236 VGKQKLYFEANGQOVKGFVTSDEGLYFVDVDSGDMWTDFTIEDKAGNWFYLGKDGAAV 1295
Db 1240 INGQHLIFRANGQOVKGFVTDHGRISYDNGSGDQIIRNFVRNAGQOVFPDNGYAV 1299
Qy 1296 TGAQTIRGOKLYPKANGQOVKGDIVKGTGDKIRYDAKSGEQVFNKTVKAAADKTYVIGN 1355
Db 1300 TGARTINGQLLYFRANGQOVKGFVTDYGRISYDNGSGDQIIRNFVRNAGQOVFPDNGYAV 1359
Qy 1356 DGVAVDPSVVGQTFKDGASGALRYFNLKQQLVTGSGWYETANHDWTYIQSGKALTCGEQTI 1415
Db 1360 NGYAV-----TGARTI 1370
Qy 1416 NGQHLIFKBDGHQVKGQVLTGDKVRYVDANSQDQAFNKSVTVNGKTYVFGNDGTAQTA 1475
Db 1371 NGQHLIFRANGQOVKGFVTDHGRISYDNGSGDQIIRNFV----- 1412

Qy 1476 GNPQQTFFDGSDIRFYSMEGQLVTGSGWYENAOQWLYV-KNGKVLTLGLQTVGSQRYVF 1534
Db 1413 -----RNAQGWYFDNNGYAVTGARTINGOHLVF 1442
Qy 1535 DENGIOAKGKAVRTSGDKIRYFDENSGSMI 1564
Db 1443 RANGVOVKGFEVTDYGRISYDANSGERV 1472

RESULT 4

US-10-383-930-34
; Sequence 34, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 34
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-34

Query Match 54.9%; Score 4580; DB 4; Length 1475;

Best Local Similarity 56.4%; Pred. No. 3.9e-242;

Matches 896; Conservative 212; Mismatches 338; Indels 144; Gaps 18;

Qy 1 MEKVRFRFMHKVKKRWVLSVASATMLASALCASASADTDTA-----SDDSNQAVVTGQ 56
Db 1 MDKVRKYLKRVKRWVSVASAVMTLTLLSGGLVKADSNESKQISNDSNTSVVTANE 60
Qy 57 TTN--NOATDPTSIATATSEQSASTDAATQASAAEQGTASTDTAAQTITNANEAK 114
Db 61 ESNVITEATSKQEAASSQTNHTVTTSSSTSVVNEKE-----VVSNPYTVGETASNGEKL- 115
Qy 115 WYPTENENGQFTDEMLAEA-----KNVATAESDSIPSD-----LAKMNVKQVDGKY 161
Db 116 -----NQTTTVTDKTSEAAANNISKQTTTEADTVIDDSNAANLQILEKLPNVKEIDGKY 169
Qy 162 YYYDQGVNKNFAYSVGDKIYFDETGAJKDTSKVDADKSSSAYSONATIFAANNRAYS 221
Db 170 YYYDNGKVRTNFTLIADGKILHFDGTAYTDSITDVTNK--DIVTTRSNLYKXNQVYD 227
Qy 222 TSAKNFEADVNTLTADSWVRPKSILKDGKWTESGKDDPRLLIMAWPDTETKRNVTNYM 281
Db 228 RQAQSEFHVHLLTAESWVRPKVILKDKTWTQSTEKDFRPLMTWVPOETQRYVNYM 287
Qy 282 NKVGVGIDKTYTATTSQADLTAAAEVLQARIEQKITSENNTKWLREASIAFVKTPQWNGE 341
Db 288 NAQLGINKTYDDTSNQLQNLAAATIAKIEAKITTLKNTDWLRQTIISAFVKVTSANSD 347
Qy 342 SEKPFDDHLQNGALLFDNOTDLTPDQSNRYLLNRTPTNQTGSLRSLRFFYNPDPLGGYD 401
Db 348 SEKPFDDHLQNGAVLYDNEGKLTTPYANSNYRILNRTPTNQTGKDPRT--ADNTIGGYE 405
Qy 402 FLLANDVDNSNPVQAEQLNWLHLLNFSGSIYANDADANFDSIRVDADVNDADLLQISS 461
Db 406 FLLANDVDNSNPVQAEQLNWLHLLNFSGSIYANDADANFDSIRVDADVNDADLLQIAG 465
Qy 462 DYLLKAAAGIDKKNKANNHVSIVEAWSDDTPLYLHDDGDNLMNMONKFLSLMWSLAKPL 521
Db 466 DYLLKAAAGIHKNDKAANDHLSILEAWSDDTPLYLHDDGDNLMNMONKFLSLFLSLAKPL 525

```
QY 522 DKXSGINPLIHNSLVBREVDREVEVTPSYSPARAHSDSEVDIIRDIKAEINPNPFGYS 581
Db 526 NORSGMPLTSLNVRDNDNAETAAPVPSYSFTRAHDDSEVDLIADIKAEINPNVVGYS 585
QY 582 FTQEEIEQAFKIYNEDLKTDKKYTHVNVPLSYTLLTNKGSIPRVYVGDMFTDDGOYMA 641
Db 586 FTWEEIKAEIYNKOLLATEKYTHYNTALSYTLLTNKSSIPRVYVGDMFTDDGOYMA 645
QY 642 NKTVNYDAIESLARKMYVSGGQAMQYQINGEILTSVRYGKGALKQSKDGADATRTS 701
Db 646 HKTINYEATLTLKARIKYVSGGQAMENQVGNSEIITSVRYGKGALKATDTGRTTTS 705
QY 702 GUGVWNGQNPFSLDGK-VVALNMGAAHQBETRALMVSTKQGVATYATDADASKAGLVK 760
Db 706 GVAVIEGNPSSLKASDRVNVNMGAAHKNQAYRPLLLTDDNGIKAYHSDQEA--AGLVR 763
QY 761 RTDENGVLFLNDLKGAVNPQVSGFLOVWVPVGAADDQDIRVAASDTASTDCK--SLHQ 818
Db 764 YNDRGELIFTAADIKGANPQVSGYLGVWVPVGA--LTKMFALRLARPHQOQMASVHQ 820
QY 819 DAAMDGRVMEGFSNFSQATKEEYTNVVIANNVDKFSWGITDFEMAPQYVSSTDGQF 878
Db 821 NAALDSRVMEGFSNFOATKKEEYTNVVIANNVDKFAEWGVTDFEMAPQYVSSTDGSR 880
QY 879 LDSVIQNGYAFTRDYDLGSKANKYGTADQLVKAIKALHAKGLKVMADWVPDQMYTFPKQ 938
Db 881 LDSVIQNGYAFTRDYDLGISKPNKYGTADDLVKAIKALHSKGIKVMADWVPDQMYAFPEK 940
QY 939 EVVTVTRTKFKGPIAGSINHSLYVTDTKSSGDDYQAKYGGAFDLDELKEKYPELTCKQ 998
Db 941 EVVTVTRTKGTPVAGSQIKNTLYVVDGKSSQOQOAKYGGAFLELOAKYPELTCKQ 1000
QY 999 ISTGQAIDPSVKIKOWSAKYFNGSNILGRGADYVLSQDVSNKYFNVA--SDTLFLPSSL 1056
Db 1001 ISTGVPMDFSVKIKOWSAKYFNGTNILGRGAGYVLDQATNYFNISDNKEINFLPKTLL 1060
QY 1057 GKVESGIRYDKGYIYNSATGQDVKASFTTEAGNLHYFGKDGVMVTGAQTINGANYFF 1116
Db 1061 NQDSQVGFSDGKGYVYVYST-SCYQAKNTFISEGDKWYTFDNNGYMVTGAQSGINGVNYFF 1119
QY 1117 LENGTAIRNTIYTDAGNSHYANCKRYENGVOQF-GNDWRYFKDGNMAVGLTIVDGNV 1175
Db 1120 LENGQLRLAILKNEGTYYVYGNDRGRRYENGYPQMSGVWRHFNNGEMSVGLTVIDGQV 1179
QY 1176 QYFEDKDGVOAKDKIIVTRDGKRYVFDQHNQNAATNFIADKTGHWYVYLGKDGVAVTGAOT 1235
Db 1180 QYFDENGYQAKGFVTTADGKIRYFDKQSGNMYRNFIEVEEGKWLILGEDGAAVTGSO 1239
QY 1236 VGKQLYFRANGQOVKGFVTSDEGLKYFYVDVSDGDMWTDFTIEDKAGNWFYLGKDGAAV 1295
Db 1240 INGQHLYFRANGVQVKGFEVTDHGRISYYDGNSGDQIRNRFVNAQGMQFYFDNNGYAV 1299
QY 1296 TGAQOTIRGOKLYFKANGQOVKGDIVKGTGDKIRYVDKSGEGBQVFNKTVKAAQKTYVION 1355
Db 1300 TGAQTINGOLLYFRANGVQVKGFEVTDYGRISYYDGNSGDQIRNRFVNAQGMQFYFDN 1359
QY 1356 DGVAVDPSVVKQOTFKDASGALRFYNLKGQLVTGSGWYETANHDMWYIQSGKALTEQITI 1415
Db 1360 NGYAV-----TGARTI 1370
QY 1416 NGQHLYFKEDGHQVKGQLVTGDKVRYDYDANGSGDQAFNKSVTVANGKTYFYFGDGTQA 1475
Db 1371 NGQHLYFRANGVQVKGFEVTDHGRISYYDGNSGDQIRNRFV----- 1412
QY 1476 GNPQKQTFKDGSDIRFYSMEGQLVTGSGWYENAOQWLYV-KNGKVLGLQTVGSGQRYVF 1534
Db 1413 -----RNAQGMQFYFDNNGYAVTGAQTINGQHLVF 1442
QY 1535 DENGIOAKGKAVRTSDGKIRYFDENSGSMI 1564
Db 1443 RANGVQVKGFEVTDYGRISYYDANGSERV 1472
```

RESULT 5

```
US-10-797-821-34
; Sequence 34, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 34
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
; US-10-797-821-34
```

Query Match 54.9%; Score 4580; DB 5; Length 1475;

Best Local Similarity 56.4%; Pred. No. 3 9e-242;

Matches 896; Conservative 212; Mismatches 338; Indels 144; Gaps 18;

```
QY 1 MEKNVRFKMHKVKRWVTLVSASATMSALGASVASADTDTA-----SDDSNQAVVTGDQ 56
Db 1 MDKVRVKLRKVKRWVTVSVASAVMTLTLSGGLVKADSNEKSQISNDSNTSVVTANE 60
QY 57 TTN--NQATDQTSIAATATSEQASASTDAATDOASAAEQGTGTTASTDTAAQTNTTANAEK 114
Db 61 ESNVTEATSKQEAASSQTNHTVTTSSSTSVNPK-----VVSNPYTVGTETASNGKL- 115
QY 115 WYPTENENOGFTDEMLAEA-----KNVATAESDSISPSD-----LAKMNVKQVDDGY 161
Db 116 -----QNTTVTDKTSAAANNISKQTTAEDTVIDDSNAANLQLEKLPNKEIDGKY 169
QY 162 YYYDQDGNVKNFVSVGDKIYYFDETGAHKDTSKYVDADKSSSAVSQNAITFAANNRAYS 221
Db 170 YYYDNGKVRTNFTLIADGKILHFDGTGAYTDTSDTVNK--DIVTTRSNLYKKYNQVYD 227
QY 222 TSANKFEAVDNYLTADSWRPKSIILKDKTWTESGKDDPRLLMAWPDOTETKRNVTNYM 281
Db 228 RSAQFEHVDHYLTAESWYRPKYILKDKGTWTQSTEKDFRLLMTWPDQETQRQYVNYM 287
QY 282 NKVVGIDKTYTAETSOADLTAAAEVLQVARIKQITSENNTKWLREASIAFVKTPQPMNGE 341
Db 288 NAQLGINKTYDTSNQLQNLTAATIQAKIEKITTLLKNTDWLRQITISAFVKTSQSAWNSD 347
QY 342 SEKPYDDHLQNGALLFDNQDITLTPDTQSNRYLLNRTPTNQTGSLDSRFTYNNPNDPLGGYD 401
Db 348 SEKPPDDHLQNGAVLYDNEGKLTTPYANSNYRLNRTPTNQTGCKDPRYT--ADNTIGGYE 405
QY 402 FLLANDVNSNPVQAEQNLNWLHYLLNFGSIYANDADANFDSIRVDADVNDVADLLQIIS 461
Db 406 FLLANDVNSNPVQAEQNLNWLHFLMNFNGIYANDPANFDSIRVDADVNDVADLLQIAG 465
QY 462 DYLKAAAGIGDKNNKNNHVSIVEAWSNDTPYLHDDGDNLMNMNKPFLSMLSLAKPL 521
Db 466 DYLKAAAGIGHKNDKKAANDHLSILEAWSNDTPYLHDDGDNMNMNKPFLSMLSLAKPL 525
QY 522 DKRSGINPLIHNSLVBREVDREVEVTPSYSPARAHSDSEVDIIRDIKAEINPNPFGYS 581
```

Db 526 NQBSGNPLITNSLVNRDNDNAETAAPVSPYSPIRAHDSVQDLADIIEKAEINPNVVGYS 585
QY 582 FTQEEIEQAFKIYNEDLKTDKYTHYNVPLSYTLTLTKNGSIPRYVYDGMFTDDQYMA 641
Db 586 FTWEEIKKAFIYNKOLLATEKKYTHYNTALSVALLTNKSVPVRYVYDGMFTDDQYMA 645
QY 642 NKTVNYDAIESLLKARMKYVSGQAMQNYQIGNEILTSVRVYKGAALKSDKGDAATTRTS 701
Db 646 HKTINYEAJETLLKARIKYVSGQAMRNOQVGNSEIITSVRVYKGAALKATDGDTRTTS 705
QY 702 GVGVMGNOPNPSLDCK--VVALNMGAAHQAQVRAALMVSTKQGVATYATDADASKAGLVK 760
Db 706 GVAVIEGNPSRLKASDRVVMNGAAHKNQAVRPLLLTDDNGIKAYHSDQEA--AGLV 763
QY 761 RTDENGILYFLNDDLKGVANPQVSGFLQVWPVPGAADDQDQIRVAASDTASTDGK--SLHQ 818
Db 764 YTNDRGELIFTAAIDIKGVANPQVSGVLGVWPVGAA--LKNFALRLARPHQOMASVHQ 820
QY 819 DAAMDVRVMEFSGNFQSFATKEEYTNVVIANNVDKFSVSGLITDFEMAPQVSSTDGQF 878
Db 821 NAALDSRVMEFSGNFQSFATKEEYTNVVIANNVDKFAEWGVTDFEMAPQVSSTDGSGF 880
QY 879 LDSVIQNGVAFTRDYDLGMSKANKYGTADOLVKAIKALHAKGLKVMADWVPDQMYTFPKQ 938
Db 881 LDSVIQNGVAFTRDYDLGMSKANKYGTADOLVKAIKALHAKGLKVMADWVPDQMYTFPEK 940
QY 939 EVVTVTRTDKFGKPIAGSQINSLYVTDTKSSGDDYQAKYGGAFPLDELKEKPELFTKKQ 998
Db 941 EVVTVTRDKYGTFAVAGSQIKNTLYVVDGKSSGKQQAQYGGAFLEELQAKYPELFPARKQ 1000
QY 999 ISTGQALDSVKIKQWSAKYFNGSNILGRGADVYVLSQVSNKYFNVA--SDTLFLPSSLL 1056
Db 1001 ISTGVPMDPSVKIKQWSAKYFNGTNILGRGAGVYVLDQATNTYFNISDNKEINFLPKTLL 1060
QY 1057 GKVVESGIRYDGGYLYNGSSATQDVKASPFITEAGNLYYFGKDGVMYVTAQTINGANYFF 1116
Db 1061 NQDSQVGSYDGGYLYYST--SGYQAKNTFISEGDKWYVFDNNGYVMTGAQSLNGVNYFF 1119
QY 1117 LENGTLALNTIYTDAGNSHYIYANDGKRYENGYQQF--GNDWRYFKDGNMAVGLTTVDGNV 1175
Db 1120 LSNGLQLRAILLKNEGDGTVAYNGDGRYENGYQQFMSGVWRHFNNGEMSVGLTVIDGQV 1179
QY 1176 QYFDDQGVQAKDIIIVTRQKVRYFDHNGNAATNTFIADKTHWYLYLKGDVAVYVTAQOT 1235
Db 1180 QYFDEMGYQAKGFVTTADGKIRYFDKQSGNMYRNFIEBEGKWLVLGEDGAATVGSQT 1239
QY 1236 VGKQKLYFEANGQVKGDFVTSDEGKLYFYVDVDSGDMWTDFTEDKAGNWFYLGKDGAAV 1295
Db 1240 INGQHLVFRANGVQVKGFEVTDHGRISYIDGNSGDQIRNRFVRNAQGVYFDNNGYAV 1299
QY 1296 TGAQTIRGQKLYFKANGQVKGDIYVKTGDKIRIYDAKSGEQVFNKTVKAAADGKTYVIGN 1355
Db 1300 TGARTINGQLLYFRANGVQVKGFEVTDYGRISYIDGNSGDQIRNRFVRNAQGVYFDNNGYAV 1359
QY 1356 DGVAVDPSVVGQTFKDGASGALRYFNKLGOLVTGSGWYETANHDWYVIOGSKALTEGQTI 1415
Db 1360 NGYAV-----TGARTI 1370
QY 1416 NGQHLVFKEDGHQVKGOLVTGDKVRYVDANSQDAFNKSVTVNGKTYFYFGNDGTAQTA 1475
Db 1371 NGQHLVFRANGVQVKGFEVTDHGRISYIDGNSGDQIRNRFV----- 1412
QY 1476 GNPKGQTFKDGDIRFYSMGBQLVTGSGWYENAGQWLYV--KNGKVLTVGLQTVGSQRVYF 1534
Db 1413 -----RNAQGVYFDNNGYAVTGARTINGQHLXYF 1442
QY 1535 DENGIOAKGKAVTSDGKIRYFDENSGSMI 1564
Db 1443 RANGVQVKGFEVTDYGRISYIDANSGERV 1472

; Sequence 4, Application US/09740274
; Patent No. US20020031826A1
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-4

Query Match 50.5%; Score 4214.5; DB 3; Length 1375;
Best Local Similarity 60.3%; Pred. No. 3.8e-222;
Matches 811; Conservative 186; Mismatches 296; Indels 53; Gaps 14;
QY 1 MEKNVRFKMHKVKKRWVTLVASATMLASALGASVASADTDTASDDSNQAVVTGDOT--T 58
Db 1 MEKKVRFKLRKVKRWVTVSIASAVVTLTSLSSLVKAD---STDDRQQAVTESQASLVT 57
QY 59 NNOATDQTSIA---ATATSEQASDAATDQASABEQTOGTTAST--DTAAQTITTNEA- 113
Db 58 TSEAAKETLTATDTSTATSATSQPTATVTDVNTNSTNQSTNTTANTANTANFVVKPTTSEQAK 117
QY 114 -----KWVPTENEN---QGFTDEM-----LAEAKNVATAESDS 143
Db 118 TNSDKKIITTSKAVNRLTATGKFPVANNNTAHPKTVTDKIVPIKPKIGKQKQSSUSQDD 177
QY 144 IPSDLAKMSNVKQDGYKYVYDQGNVKNFVSVGDKIYVYFDETGAYKDTSKVDADKSS 203
Db 178 IAA-LGNVKNIRKVGKYYKYKEDGTLQKNYALNKGKTFPFDGTCALSNNLT-LPSKKN 235
QY 204 SAVSQNATIFAANNRAYTSKAFNFEAVDNYLTADSWYRPKSILKDGKTWTESGKDDPRPL 263
Db 236 ITNNDNTNSFAQVNOVYSTDVANFEHVDHYLTAEWSYRPKYILKDGKWTQTSTQTEKDFRPL 295
QY 264 LMAWPDPTETKRYNNYNNKVVIGDKTYTAETFSQADLTAAAEVLQARIKQITSENNTKW 323
Db 296 LMTWMPDQETQRYNNYNNALQIHQTYNTATSPLOLNLAQAQTIQTKIBEKITAENKNTW 355
QY 324 LREASAFVKTOPWNGESEKPYDDHLQNGALLFDNQDTDLTPDTQSNRYLLNRTPTNQGT 383
Db 356 LRQTTISAFVKTOSAWNSDSEKPFDDHLQKGLLYSNNKLTQANSNYRILNRTPTNQGT 415
QY 384 SLDSRTYNNPNDPLGQYDFLLANDVDNSPNVQAEQLNMHLHYLLNFGSIYANDADANFDS 443
Db 416 KKDPRTY--ADRTIGGYEFELANDVDNSPNVQAEQLNMHLFLMNFNGIYANDPDANFDS 473
QY 444 IRVDVNDVNDADLQISSDYLKAAVYGIKNNKNNHVSIVEAWSNDPTPLYHDDGDNLN 503
Db 474 IRVDVNDVNDADLLQIAGDYLKAAKGHIHKNDKAANDHLSLLEAWSYNDPTPLYHDDGDNMI 533
QY 504 NMDNKPRLSMLSLAKPLDKRSGINPLIHNLSVDREVDDREVETVPYSFARAFHDSVQD 563
Db 534 NMDNRLRLSLYSLAKPLNQSRGMPNPLITNSLVNRDNDNAETAAPVSPYSPIRAHDSVQD 593

RESULT 10
 US-10-383-930-36
 ; Sequence 36, Application US/10383930
 ; Publication No. US20040127400A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Smith, Daniel J
 ; APPLICANT: Taubman, Martin A
 ; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
 ; FILE REFERENCE: 25669-018
 ; CURRENT APPLICATION NUMBER: US/10/383,930
 ; CURRENT FILING DATE: 2003-03-07
 ; PRIOR APPLICATION NUMBER: 60/402,483
 ; PRIOR FILING DATE: 2002-08-08
 ; PRIOR APPLICATION NUMBER: 60/363,209
 ; PRIOR FILING DATE: 2002-03-07
 ; NUMBER OF SEQ ID NOS: 41
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 36
 ; LENGTH: 1430
 ; TYPE: PRT
 ; ORGANISM: Streptococcus mutans
 US-10-383-930-36
 Query Match 42.8%; Score 3575; DB 4; Length 1430;
 Best Local Similarity 48.4%; Pred. No. 4.5e-187;
 Matches 719; Conservative 240; Mismatches 408; Indels 118; Gaps 26;
 Qy 1 MEKNVRFKMHKVRWVTLSSVASA--TMLASALGASVAS-----ADTDAS 44
 Db 1 METKRRYKMHKVKHWWVAVASGLITLGTTLIGSSVSAETEQTSDKVVTKQSEDDKAA 60
 Qy 45 DDSNQA---VVGDTNNQATQTSIAATATS-EQASSTDAATDOASAAEQGTGTAST 100
 Db 61 SSSQTDAPKTKQAQTEQTOAQSQANVADTSTSIKETPSSQNIITQANSDDKTVNTKSE 120
 Qy 101 D--TAAQTNTTNAEAKWPTENENQGFDEMLAEAKNVAT-AESDSIPSDLAKMSNVKOV 157
 Db 121 EAQTSERTKQSEAQ---TTASSQALTQAKBLTKQRTAQENKKNPVDLAAIPNVKQI 177
 Qy 158 DGKYYYDQDGNKKNFPAVSGDKIYYFDE-TGAYKDTSKVDADKSSASVSQNAITFAAN 216
 Db 178 DGKYYYIGSDGPKKNFALTVNNKLVLYFDKNTGALTDTSTQYQPKQGLTKLND---YTPH 234
 Qy 217 NRAYSTAKNFEAVDNVLTADSWYRPKSILKQKWTWESGKODFRPLMAWHPDPTETKRN 276
 Db 235 NQIVNFENTSLKETIDNVTYVADSWYRPRKDLKNGKWTWTASSESLRPLLLMSWMPDKQTQTA 294
 Qy 277 VYNNMKV-VGIDKTYTAETSQADLTAAAEVLQARIEOKITSENNTKWLREALSAPVKTO 335
 Db 295 YLNTMNOQGLTGENTYADSSQESLNLAQTVQVKLETKISQTOQTOWLRDINSFVKQ 354
 Qy 336 PQWGESEKPYD---DHLQNGALLFDNQDQLTDFDQSNRYLLNRTPTNTQTSLSRFTY 391
 Db 355 PNWNSQTESDTSAGEKDHQGGALLYSN-SDKTAVANSYRLLNRTPTSTQTG---KPKY 409
 Qy 392 NPNDPLGGYDFLLANDVNDNSNPVQAEQLNWLHYLNFSGSIYANDADANFDSIRVDAVN 451
 Db 410 FEDNSGGYDFLLANDINDSNPVQAEQLNWLHYLNNYSGSIYANDPEANFDGVRVDAVN 469
 Qy 452 VDADLLQISDYLKAAAGIDKNNKNNHVSIVEAWSNDNTPVLDHGDGNLMMNDKPL 511
 Db 470 VNADLLQISDYLKAHGVGVDSKKNALNHLISLEANSDDNDPQYKDTKGAQPIPNKLR 529
 Qy 512 SMLWSLAKPLDK-----RSLGNPLNHLNSVDREVEVTPSVSFARAHSEVQD 563
 Db 530 SLLVALTRPLEKDAASKNKEIRSGLEPVITNSLNNRSGEKNRSMANYIFIRAHSEVQT 589
 Qy 564 IIRDIKAEINPNPSFGVSFTQBEIQAFKLYNEDELKTKKYHYNVPLSYTLTLTNKGS 623
 Db 590 VIAKIIKQAINPKPTDGLTLDLQKAFKLYNEDEMRQAKKKYQTSNIPTAYALMLSNKDS 649
 Qy 624 IPVYYGDMFTDGOYWAFTKSPYDALESLLKARMKYVSGGOAMO-NYOINGE----- 676

;; PRIOR APPLICATION NUMBER: 09/290,049
;; PRIOR FILING DATE: 1999-04-12
;; PRIOR APPLICATION NUMBER: 60/081,550
;; PRIOR FILING DATE: 1998-04-13
;; PRIOR APPLICATION NUMBER: 60/115,142
;; PRIOR FILING DATE: 1999-01-08
;; NUMBER OF SEQ ID NOS: 45
;; SOFTWARE: PatentIn version 3.2
;; SEQ ID NO 36
;; LENGTH: 1430
;; TYPE: PR1
;; ORGANISM: Streptococcus mutans
US-10-797-821-36

Query Match 42.8%; Score 3575; DB 5; Length 1430;
Best Local Similarity 48.4%; Pred. No. 4.5e-187;
Matches 719; Conservative 240; Mismatches 408; Indels 118; Gaps 26;

QY 1 MEKNVREKMKVKRWVTLVSASA--TWLASALCASVAS-----ADDTAS 44
DB 1 MEKRRYKMKVKRWVTVAVASGLITLGTTLGSSVSAETEQQTQSKVVTQKSEDDKAA 60
QY 45 DDSNQA---VVTGDTNNQATDTSIAATATS--EQSASTDAATDOASAAEQTGTTAST 100
DB 61 SESQTDAPTKQAQTEQTQAQSQANVADTSTSTKETPSQNTTQANSDDKTVNTKSE 120
QY 101 D--TAAQTNNANEAKWVTENENQGFTEMLAEAKNVAT--AESDSIPSLAKMSNVKQV 157
DB 121 EAQTSERTKQSEBAQ---TTASSQALTOAKAELTKQRTAAQENKPNVDLAALPNVKQI 177
QY 158 DGKYYVYDQGNVKKPANSVGDKIYYFDE--TGAYKDTSKVDADKSSASVQNAIFAA 216
DB 178 DGKYYVYIGSQPKQNFALVNNKVLVFDKNTGALTDTSQYQFKQGLTKLND---YTPH 234
QY 217 NRAYSTSANKFEAVDNYLTADSNVRPKSILKCKGKWTESGKDFRPLLMAMWPDTEKRN 276
DB 235 NQIVNFTSLETIDNVTADSNVRPKDILKNGKWTWASSESDLRLPLMSWFPDKQTQIA 294
QY 277 YVNNMKV--VGIDKTYTAETSQADLTAAAEVLQVARIQKITSNNTKWLREALSAFVKQT 335
DB 295 YLNNMQOGLGTGENTYADSSQSLNLAQTVQVKIETKISQTOQOWLEDLINSFVKQT 354
QY 336 PQWNGSESEKPYD---DHQNGALLFDNQDLDTPDQSNVRLNRTPTNQTGSLDSRFTY 391
DB 355 PNWNSQTESDTSAGEKDHQGGALLYSN--SDKTAYANSVRYLLNRTPTSTQ---KPKY 409
QY 392 NPNDPLGGYDFLLANDVNSNPVVQAEOLNWLHLLNFGSIYANDADANFDSIRVDADV 451
DB 410 FEDNSSGGYDFLLANDIDNSNPVVQAEOLNWLHLLNFGSIYANDPEANFDGVRVDADV 469
QY 452 VDADLLQISDYLKAAAYGIDKNNKANNHVSIVEAWSNDNTPYLHDDGDNLMNMNDKFL 511
DB 470 VNADLLQIASDYLKAAHYGVDSKSEKAINHLISLEAWSNDNTPQNTKTKGAQLFDNKLRL 529
QY 512 SMLWSLAKPLDK-----RGLNPLHNSLVDRVEDVETVPSYSFARAHDSVQD 563
DB 530 SLLYALTRPLEKDAOSKNEIRSGLEPVITNSLNNRSAGKNSERMANYIFIRAHDSVQT 589
QY 564 IIRDIILKAINPNSFGYSTOEIEQAFKYNEDLKKTKYTHYVPLSYLLTLNKG 623
DB 590 VIATIKAIQINPKTDGLTFLDELKQAFKIYNEDMRQAKKTKYTSNIPTAYALMLSNKDS 649
QY 624 IPRVYVYDGMFTDDQGYMANKTVNYDAIESLLKARMKYVSGQAWQ--NYQINGE----- 676
DB 650 ITRLYYGDVYSDGQYMAKSPYDAIDTLKARIKAYAGGQDMKITYVEGDKSHMDWY 709
QY 677 --ILTSVRYGKALKQDGDATRTTSVGVVGNQPNFSLD--GKVALNMGAAHANQRY 733
DB 710 TGVLTSVRYGTGANEATDQSEATKTKQGMVITSNNPVLKLNQNDKVIYVNMGAHKNQRY 769
QY 734 RALMVSTKDVATYATDADAKAGLVKRTDENGILYFLNDDLKGVANPOVSGFLQVWVPV 793
DB 770 RPLLTTKDGLTSYTSDAAK--SLYRKTNDKGLVFDASDIQGYLNPQVSGYLAVWVPV 827

QY 794 GAADDQDIRVAASDTASTDGKSLHQDAAMDSRVMEFEGFSFQSPATKEEYTVNNVIANNV 853
DB 828 GASDNQDVRVAASNKANATQVYESSALDSQLIYEGFSNPDQFVTKDSDTNNKIAQNV 887
QY 854 DKFVSGWITDFEMAPQVSVSTGDFLDSVITQNGVAFTRDYDLGMSKANKYGTADQLVKAI 913
DB 888 QLFKSGWVTSFEMAPQVSVSSEDSGLDSIIQNGVAFEDRYDLAMSNNKYSQQDMINAV 947
QY 914 KALHAKGLKVMADWVPDQMYTFFKQEVVTVTRTDKFKGKPIAGSINHSIYVTDTKSSGDD 973
DB 948 KALHKSIGIQAIDWVPDQIYNLPKGEVVTATRVNDYGEYKDSSEIKNTLYAANTKSNKGD 1007
QY 974 YQAKYGGAFDELKEKYPELFTKKQISTGOAIDPSVKIKOWSAKYFNGSNILGRGADYVL 1033
DB 1008 YQAKYGGAFSELAELAAKYPISFNRTQISNGKKIDPSEKITAOKAKYFNGTILGRGVYVL 1067
QY 1034 SDQVSNKYFNVAADTLPLSSLLGKVVESGIRVDGKGYIYNSSATGQOVKASFITEA--GN 1092
DB 1068 KDNASDKYFELKGNQTYLPRQMTNKEASTGFVNDGNGMTFYST--SGYQAKNSFVQDAKGN 1126
QY 1093 LYTFEGDKGYMVTGAQTINGANYFFLENGTALRNTIYTDAGNSHYIYANDGKRYENGQOQ 1152
DB 1127 WYTFDNNGHVYGLQOLNGEVQYFLSNGVQLRESFLENADGSKNYFGLHNGNYSNGYYSF 1186
QY 1153 GND--WRYF--KDCGNMAVGLTVDGNVQYFDKGVQAKDKIIVTRDGVRYFDQHNNAAT 1209
DB 1187 DNDKWRYPDASGVMAVGLKTINGNTQYFDQDGVQVKGAWITGSDGKKRFPDDSGNNAV 1246
QY 1210 NTFIADKTHWYLLGKDGVAVTGAQTVGKQKLYFEANGQQVKGDFVTSDEGLKYFYDQVDS 1269
DB 1247 NRPANDKNGDWYLLNSDGIALGVQTINGKTYVFGQDGKQIKGKIIT--DNGKLYFLANS 1305
QY 1270 GDMWTDFTEDKAGNWFYLGKDAATGAQTIRGQKLYFKANGQQVKGDIIVKGTGDKIRY 1329
DB 1306 GELARNIFATDSQNNWYFSGDGVAVTGSQTIAGKLYFASDGKQVKGSEFT--YNGKVHY 1364
QY 1330 YDAKSQGEQVFNKTVKAADGKTYVIGNDGVAVDPSVVKGTFFKDGASGALRFVNLKGQVLTG 1389
DB 1365 YHADSSELQVNRPEADKDG----- 1393
QY 1390 SGWYETANHDWVYIQS--GKALTGEQTINGQHLFYKEDGHQVKGQ 1433
DB 1384 -----NWYLDNSGEALTGSORINDQVFTTREGKQVKGDV 1419

RESULT 12
US-10-383-930-38
; Sequence 38, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 38
; LENGTH: 1554
; TYPE: PR1
; ORGANISM: Streptococcus sobrinus
US-10-383-930-38

Query Match 42.0%; Score 3502.5; DB 4; Length 1554;
Best Local Similarity 45.4%; Pred. No. 4.7e-183;
Matches 749; Conservative 252; Mismatches 482; Indels 167; Gaps 40;

Db 60 SLVTSPEATKEADKRTNTEADVLTPAKETNAVETATTTNTQATAEAAATATTADAVAA 119
Qy 91 EOTQGTASTDTAAQTITNANE-AKWPTENENOGFTDEMLAEAKNVAETAESDISPSDLA 149
Db 120 VPKAEAVTTTDAVAVTKEAEQPAATVKAENV-----TEVKAPEALKDSEVEAL 171
Qy 150 KMSNVQVDKYYYYDODGNVKNFVSGDKIYYFDETCAYKXDTSKVDADKSSVASQON 209
Db 172 SLKNIKNIDKYYVNEDESGHKNFALTNGQLYFGKOGALTSSSTYSFTPGTINVDG 231
Qy 210 ATIFAAANRAYSTAKNFEAVDNYLTADSWYRPSILKDGKWTESKODFRLLMAWPP 269
Db 232 ---FSINRAYDSEASEFELIDGLTADSWYRPSIIKDGTWOASTAEDFRLLMAWPP 288
Qy 270 DTETKKNYNNMKNVGVIDKTYTAETSQADLTAAAEVLQARIQKITSENNTKWLREAS 329
Db 289 NVDTOVNLNYSKVNFLDAKYSSDTQETLKVAAKDIQIKIEQKQAEKSTQWLREAS 348
Qy 330 AFVKTOPOWNGESEKPY-----DHLQNGALLFDNQTDLTPDTQSNRYLLNRTPNTQGS 384
Db 349 AFVKTOPOWNGESEN-YSKGGGDLQGGALLVND-SRTPWANSYRRLNRTPNTQGT 406
Qy 385 LDRSFTYNDP--LGSYDFLANDVNSNPVQAEQLNWLHLLNFGSIYANDADANED 442
Db 407 IDKSLDEQSDPNHMGDFLLANDVLSNPVQAEQLNWLHLLNFGSIYANDADANED 466
Qy 443 SIRDAVDNVDALLQISSDYKAAAYGIDKNNKNANNHVSIVAEWSDNDTPYLHDDGDNL 502
Db 467 GIRDAVDNVDADMLQLYTNYFREYGVNKNSEANALAHISVLEANSINDNHNKDTDGAA 526
Qy 503 MWNNDKFRSLMWSLAKPLDKGS-GLNPLHNSLVDRVDDR-----EVEV 548
Db 527 LAMENKQRLALLFLAKPIKERTPAVSPLYNNTFNTTQDEKTDWKNGSKAYNEDGT 586
Qy 549 P-----SYSEARHDSVODIIRDIKAEINPNSGYSFTQEEIEQAF 592
Db 587 KOSTIGKYNKYGASGNVFRADHNNVODIAEIIKKEINPKSDGFTITDAEMKQAFE 646
Qy 593 IYNEDLKTDKKTTHYNPLSYLLTNKGSIPRVYVGMFTDDGQYMANKTWNYDAIES 652
Db 647 IYNKMLSSDKKYLNNIPAAVAVMLQNMETITRVYVGLYDLDGHYMETKSPYDTIYN 706
Qy 653 LKARKKYYGSGOAMQYQI-----GNGEILTSVYKGGALKQSD-KGDAT 697
Db 707 LMKSRIKYVSGGOAQRSYWLPDGMKNSDVELYRTNEVYTSVRYGKIDMTANDTEGSKY 766
Qy 698 TTSVGVMGNQPNFSLDGKV-VALMGAHANQYRALMVSTKDGVAATYATDADASKA 756
Db 767 SRTSGOVLVANNPKLNLDSQAKLNVEMGKIHANQYRALIVGTADGINKFTSDADAI 826
Qy 757 GLVKRTDENGILYFLNDDLKGVANPOVSGFLOVVPVGAADDODIRVAASDTASTDGK-S 815
Db 827 GYVKETDSNGLVTFGANDIKGYETDMSGFVAVWVPVGSADNODIRVAPSTAKKEGELT 886
Qy 816 LHQDAAMPDSRVNFEFSNQSF--ATKEBEYTNVVIANNVDKFSWNGITDFEMAPQYVSS 873
Db 887 LKATEAYDSQLIYEGFSNFQITPDGSDPSVYNNRKAENVDLFKSGVTSFEMAPQYVSA 946
Qy 874 TDQQLDSVIQNGYAFTRDYLGMKANKYGTADQLVKAIKALHAKGLKVMADWVPDQMY 933
Db 947 DDGTFELDSVIQNGYAFADRYDLAMSKNNKYSKEDLDALKALHAKGIAQIAIADWVPDQY 1006
Qy 934 TPKQGVVTVTRTKFPGIAGSOINHSLYVTDTKSSGDDYQAKYGGAFDELKEKYP 993
Db 1007 QLPGEVTVTRTKDAGRKIADAIIDHSLYVANSKSSGKYQAKYGGEFELAEKAKY 1066
Qy 994 FTKQISTQQAIDPSVKIKQWAKYFNGSNIILGRGADYVLSQVSNKYFNVASDTLFLPS 1053
Db 1067 FKVNIMSTGKPIDDSVKLQWAKYFNGTNVLERGVYVLSDBATCKYFTVTKGNFPL 1126
Qy 1054 SLLGK-VVESGIRYKGVYNNSSATGQVKASFITEAGNLVYFGKGVMTGAQ-TING 1111
Db 1127 QLTGKEVITGFSDDGKITTY-FGTSGTQAKSAFVTFNGNTYTFDARGHMTNSBSPNG 1185

Qy 1112 AN-YFLENGTALRNTIYTDACGNSHYANDGKRYENGYOOF-----GND-----WR 1157
Db 1186 KDYRFLPGLMISLNAFYDANGNTYLYNSKGQMYKGYTKFDVSETDKDKGSKVKVFR 1245
Qy 1158 YF-KDGNMAGVLTVDNNGVQYFDKDGVOAKDKIIVTRDGRKRYRFDQHNGNAATNTFIADK 1216
Db 1246 YFTNEGVMAGVTVIDGFTQYFGEDEGFOAKDK-LVTFKGYTYFDAHTGNGIKDTW-RNI 1303
Qy 1217 TGHWYLLGKDGVAVTGAQTVGKOKLYFEANGQOVKGDFTVSDGKLYFYDVSDDMWDT 1276
Db 1304 NGKWIYFDANGVAATGAQVINGOKLYFNEDEGSQVKGWKNADGTYSKYKEGFEELVNE 1363
Qy 1277 FIEDKAGNWFYLGKDGAAVTGAQTVGKOKLYFKANGQOVKGDIVKGTGKIRYVDAKSGE 1336
Db 1364 FFTTDGNWYVYAGANKTVTGAQVINGQHLVFNADSGOVKGGVVKADGTYSKYNASTGE 1423
Qy 1337 QVENKTVKAADGKTYVYVIGNDGVAVDPVVKGTFFDASGALRFYNLKGLVGTSGWYETA 1396
Db 1424 RLTVNE-----PFTTG 1433
Qy 1397 NHDWYI-OSGKALTGEQTINGQHLVFKEDGHQVKGOLVGTGDKVRYVDANSQDQPNK 1455
Db 1434 DNNWYVIGANGKSVTGEVKIGDDTYFFAKDGKQVKGQTVSAGNGRISYYYGDSGKRAVST 1493
Qy 1456 SVTVNGKTY-YFGNDGTA 1472
Db 1494 WIEIQGVTVYVFDKNGLA 1511

Search completed: February 11, 2006, 20:57:01

Job time : 151.315 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 20:42:47 ; Search time 9.04803 Seconds
(without alignments)
2306.008 Million cell updates/sec

Title: US-10-797-821-37
Perfect score: 8349
Sequence: 1 MEKNVRFQHKVKKRWTLIS.....VYQYYVFGDGRYRGWN 1590

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 97014 seqs, 13122538 residues

Total number of hits satisfying chosen parameters: 97014

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA_New.*
1: /cgn2_6/prodata1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/prodata1/pubpaa/US06_NEW_PUB.pep.*
3: /cgn2_6/prodata1/pubpaa/US07_NEW_PUB.pep.*
4: /cgn2_6/prodata1/pubpaa/PCT_NEW_PUB.pep.*
5: /cgn2_6/prodata1/pubpaa/US05_NEW_PUB.pep.*
6: /cgn2_6/prodata1/pubpaa/US10_NEW_PUB.pep.*
7: /cgn2_6/prodata1/pubpaa/US11_NEW_PUB.pep.*
8: /cgn2_6/prodata1/pubpaa/US60_NEW_PUB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	513	6.1	2710	7	US-11-051-453-41
2	443.5	5.3	726	7	US-11-052-554A-217
3	381	4.6	2367	7	US-11-051-453-42
4	285.5	3.4	2314	7	US-11-013-759-11
5	284.5	3.4	2053	7	US-11-013-759-9
6	279.5	3.3	1992	7	US-11-013-759-3
7	279.5	3.3	1992	7	US-11-013-759-13
8	279.5	3.3	2047	7	US-11-013-759-4
9	279.5	3.3	2047	7	US-11-013-759-7
10	265	3.2	1588	7	US-11-052-554A-280
11	258.5	3.1	5024	6	US-10-793-626-2364
12	257	3.1	396	7	US-11-022-562-228
13	254.5	3.0	658	6	US-10-873-528-17
14	254.5	3.0	677	6	US-10-873-528-155
15	249.5	3.0	1155	6	US-10-793-626-1780
16	248	3.0	619	7	US-11-052-554A-229
17	242.5	2.9	3194	7	US-11-052-554A-90
18	241.5	2.9	5291	7	US-11-052-554A-281
19	236.5	2.8	690	7	US-11-052-554A-232
20	234	2.8	2902	7	US-11-052-554A-91
21	228	2.7	1448	6	US-10-485-517-212
22	226	2.7	701	7	US-11-052-554A-231
23	225	2.7	744	6	US-10-873-528-184
24	223.5	2.7	1647	7	US-11-052-554A-260
25	214	2.6	338	7	US-11-052-554A-228

26	212	2.5	2399	7	US-11-052-554A-92	Sequence 92, Appl
27	211.5	2.5	1461	7	US-11-052-554A-283	Sequence 283, App
28	209	2.5	483	7	US-11-113-775A-1	Sequence 1, Appli
29	209	2.5	483	7	US-11-113-799-1	Sequence 2, Appli
30	209	2.5	1571	7	US-11-052-554A-2	Sequence 2, Appli
31	207.5	2.5	501	6	US-10-830-203-25	Sequence 25, Appl
32	207	2.5	1562	7	US-11-052-554A-211	Sequence 211, App
33	206	2.5	483	7	US-11-113-775A-2	Sequence 2, Appli
34	206	2.5	483	7	US-11-113-799-2	Sequence 2, Appli
35	203	2.4	980	7	US-11-052-554A-17	Sequence 17, Appl
36	202	2.4	1767	7	US-11-052-554A-372	Sequence 372, App
37	201	2.4	485	6	US-10-630-203-12	Sequence 12, Appl
38	201	2.4	485	7	US-11-113-775A-3	Sequence 3, Appli
39	201	2.4	485	7	US-11-113-799-3	Sequence 3, Appli
40	199	2.4	2340	7	US-11-052-554A-171	Sequence 171, App
41	198	2.4	1345	7	US-11-052-554A-282	Sequence 282, App
42	197	2.4	1263	6	US-10-485-517-127	Sequence 127, App
43	196	2.3	1107	6	US-10-485-517-145	Sequence 145, App
44	195	2.3	501	6	US-10-630-203-27	Sequence 27, Appl
45	195	2.3	501	6	US-10-630-203-28	Sequence 28, Appl

ALIGNMENTS

RESULT 1

US-11-051-453-41
; Sequence 41, Application US/11051453
; Publication No. US20050287150A1
; GENERAL INFORMATION:
; APPLICANT: AMBROSINO, DONNA
; APPLICANT: BABCOCK, GREGORY J.
; APPLICANT: BROERING, THERESA
; APPLICANT: GRAZIANO, ROBERT
; APPLICANT: HERNANDEZ, HECTOR JAVIER
; APPLICANT: LOWY, ISRAEL
; APPLICANT: MANDELL, ROBERT
; APPLICANT: MOLRINE, DEBORAH
; APPLICANT: THOMAS, JR., WILLIAM D.
; APPLICANT: ZHANG, HUI-FEN
; TITLE OF INVENTION: ANTIBODIES AGAINST CLOSTRIDIUM DIFFICILE TOXINS AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: MJ1-001
; CURRENT APPLICATION NUMBER: US/11/051,453
; PRIOR FILING DATE: 2005-02-04
; PRIOR APPLICATION NUMBER: 60/542,357
; PRIOR FILING DATE: 2004-02-06
; PRIOR APPLICATION NUMBER: 60/613,854
; PRIOR FILING DATE: 2004-09-28
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 41
; LENGTH: 2710
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-11-051-453-41

Query Match 6.1%; Score 513; DB 7; Length 2710;
Best Local Similarity 21.2%; Pred. No. 6.9e-20;
Matches 364; Conservative 242; Mismatches 619; Indels 496; Gaps 92;

QY	119	ENENQFTDE--MLAEA-KNVATAESDSTPSDLAKMSN--VKQVD-----GKYY----	162
Db	1211	ETENLDFSKIMLMNAPSRSVFWETGAVPG-LRSLNDGTRLLDSIRLDYFGKFTWRFY	1269
QY	163	YYD-----QDGNVK-----KNFAV-----SVGDKI-YYFDGTG-----	189
Db	1270	AFYDVAITTLKPVYEDTNIKIKLDKXDRNFIMPTTNEIRNKLSYSPDGAGTYSLLS	1329
QY	190	AYKOTSQVDADKSSSAVSONATIPANNRAYTSKAFVADNYLTADSWYR-----PKSI	245
Db	1330	SYPISTINLSK-----DDLWIFNDREVISIENGTKKGKLIKVDLSKIDINKNKL	1383

Db 198 SDGNKHYYDRDSEMTNRFFVNDQ--GNWYYLNDGVVPTGSIITVNGOSLYFNSDGSQVKG 256
QY 1318 DIVKGTGKIRYYDAKSGEQVFNKTKAADGKTYVIGNDG--VAVDP-SYVKGQTFKDA- 1373
Db 257 NFVE-EDGSLRYDKNSGD--LLRKTSTRTINGVNYQFNDGNGARAIKIEVVKTSLVUDSY 314
QY 1374 -----SGALRFYNLK-----GQVLTSGHWYETANHDWVYIQSGKALTGSEQTNGQH 1419
Db 315 EFGPSVSKILFBNHKKVTPAVVHAGAMVTTAGVQRKILNSYVSNASG-----H 362
QY 1420 LYFKEDGHQVKGOLVGTGDKVRYDYDANSQDAFNKSVTVNGKTYVFGNDGTAQTAGNPK 1479
Db 363 VTFDSSHVYTLSDIP-----YDPN--DSSRNASPFIFDSPAARNWVNSYTVKVDN 413
QY 1480 GQTFKDGSD 1488
Db 414 LQVQADGSN 422
RESULT 3
; Sequence 42, Application US/11051453
; Publication No. US20050287150A1
; GENERAL INFORMATION:
; APPLICANT: AMBROSINO, DONNA
; APPLICANT: BABCOCK, GREGORY J.
; APPLICANT: BROERING, THERESA
; APPLICANT: GRAZIANO, ROBERT
; APPLICANT: HERNANDEZ, HECTOR JAVIER
; APPLICANT: LOWY, ISRAEL
; APPLICANT: MANDELL, ROBERT
; APPLICANT: MOLRINE, DEBORAH
; APPLICANT: THOMAS, JR., WILLIAM D.
; APPLICANT: ZHANG, HUI-PEN
; TITLE OF INVENTION: ANTIBODIES AGAINST CLOSTRIDIUM DIFFICILE TOXINS AND
; FILE REFERENCE: MJI-001
; CURRENT APPLICATION NUMBER: US/11/051,453
; PRIORITY FILING DATE: 2005-02-04
; PRIOR APPLICATION NUMBER: 60/542,357
; PRIOR FILING DATE: 2004-02-06
; PRIOR APPLICATION NUMBER: 60/613,854
; PRIOR FILING DATE: 2004-09-28
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 42
; LENGTH: 2367
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-11-051-453-42
Query Match 4.6%; Score 381; DB 7; Length 2367;
Best Local Similarity 21.1%; Pred. No. 8.3e-13;
Matches 292; Conservative 203; Mismatches 474; Indels 412; Gaps 82;
QY 356 LFDNOTDLPTDQTS-----NVLNRPTPTNOTGSLDSRF-----TVPNDPLGGY- 400
Db 1230 VFAMEGTGTPGLSLENDGTKLLDRIDNVEGEFYWRYFAFTADALITLKR-----YE 1284
QY 401 DFLANDVDNSN-----PVQA-----EQLNLHYLLNFGSIYA-----NDADANFD 442
Db 1285 DTNIRINLDSNRSFVPIITTEYIREKLSYSPY--GSGGTVALPLSQYMMGINIELSES 1342
QY 443 STRVAVDNDVADLLOISDYLK-----AAYGIDKNNKNNHNSIVEAWSNDTP 493
Db 1343 DWIITDIDVNVNRD-VTIESDKTKKGLIEGILSTLSIEENKILNSH----- 1388
QY 494 YLHDDGNLMMNDKFRLSMLAKPLDKRSGLNPLIHNSLVDREVDDREVETVPSYF 553
Db 1389 EINFSGE--VNSNGF-VSLTFSILE-----GINAIEVDLSK-----SYKL 1428
QY 554 ABRAHSEVDIIRDIKAEINPNSFGYSFTQBEIEQAFKIYNEDLKKTKDKYTHNVPLS 613

Db 1429 -----LISGELKILMLNSNHQQKID--YIGFNSBELQK-----NIPYS 1464
QY 614 YTELLLTNKG-----IPRVYVYGDMTFDDCOYMAN-----KTVN 646
Db 1465 F--VNSEKENGFIINGSTKEGLFVSELDPVVLISKVYMDDSKPSFGYVNNLKVQVIT 1521
QY 647 YDAIESL-----LKARMKYVSGGQAMQNY-----QIGNGEILTSTVRYGKALKQS 691
Db 1522 KQNVNLTGYLKDDIK-ISLSLTLQDEKTIKLSVHLDSEGVAEILKFWN-RKGSTNWS 1579
QY 692 DKGDATRTTSGVGVWGN--OPN--FSLDKVVALNMGAAHQAQVPRALMVSTKDGVTATY 747
Db 1580 DSLMSFLESNWKISIFVNFLOSNIKFILDANF-----ISGTTISGQFEP 1624
QY 748 ATDADAS-KAGLVK-RTDENGILYFLNDDLKGVANQVSGFLQVWVPVGAADQDQIRVAA 805
Db 1625 ICDENNNIOPYFIKFTNLTETNYLYVGNRQNMIVEPYD-----LDDSGDI---S 1671
QY 806 SDTASTDGKSLHODAAAMDSRV-----MFEFGSNFQSPATKEEBEYTNVI--ANNVDK 855
Db 1672 STVINFSQKLY--GIDSCVNKVVISPNIIYDEINITPVYETNNTYPEVILVDANYINE 1728
QY 856 FVSWGITDFEMAPQVYSSTDGQ-----FLDSVIQN--GVAPT 890
Db 1729 KINVNINDLSI--RYVMSNDGNDFILMSTSEENKVSQKIRFVNVFKDKTLANKLSFNFS 1786
QY 891 DRYDLGMSKANKYGTADQVKAIKALHAKGL-----KVMADWVPDQVTPPKQSVTVTRT 946
Db 1787 DKQDVPVSEI-----ILSFTPSYVEDGLIGYDLGLVSLYNEKY-----I 1826
QY 947 DRFGKPIAG-SQINHSLY-----VTDTKSSGDDYQAKY-----GGAFLDELKBY 990
Db 1827 NNFQGMVSGLIYINDSLYFKEPVNLIITGFTVGGD--KYFNPINGCA----- 1874
QY 991 PELFTYKQISTQOAIIDPSVKIKQWSAKYFNGSNILGRGADYVLSQVSNKYFNVASDTLP 1050
Db 1875 -----ASIGETI---IDDKNY---YFNQSGVLQTG--VFSTEDGKFYFAPANT--- 1914
QY 1051 LPSLLGKVVESGIRYDGHGYLYNSSATGDQVKASFIETAGNLYYFGKQGYMTGA--Q 1107
Db 1915 LDENLEGEAID---FTGKLI---DE-----NIYF-EDNY--RGAVEWK 1950
QY 1108 TING-ANYFFLENGTALR--NTIYTDAQNSHYIANDGKRYENGYOQFGNDMRYFKD-GN 1163
Db 1951 ELDDGEMHYFSPETGKAFGLNLI-----GDDKYFNSDGMQKGFVINDNKHVFDSDGV 2005
QY 1164 MAVGLTTVDGNVOYFDKQVQAKKIIIVTRDQKRVYFDQHNHNAATNTFIADKTGHYYL 1223
Db 2006 MKVGYTEIDGKHIFYAENG-EMQIGVFNTEDG-FKYFAHHNED-----LGNEEG----- 2052
QY 1224 GKDGVAVTGAQTVGKQKLYFEANGQOVKGFVTSDEGKLYFYVDVSDGMWTD-TFIEDKA 1282
Db 2053 ---EESISYGLNPNKNIYYFDDSFATAVG-WKLEDGSKYFFPEDTAEAYIGUSLIND-- 2107
QY 1283 GNVFYLKDGGAATGAQTIRGQKLYFKANGQOVKGDIVKGTGDKIRIYDAKSGEQVFNKT 1342
Db 2108 -QOYFNDGDIQVGFVTVINDKVFYFSDSGLIESG--VQNIIDNIFYIDNGGIVQI--GV 2162
QY 1343 VKAAGKTYVIG-----NDGVAVDPSVVKQGTQFKDASGLR-----FYNLKGQVLTSGSW-Y 1393
Db 2163 FQTSQGYKYFAPANTVNDN-----IYQAV-EYSGLVRVGDVYFYFETTYIETGIWY 2214
QY 1394 ETANHDWVI---QSGKALTGQTINGQHLKPKEDGHQVKGQVLTG--IDGKRVYVDANS 1448
Db 2215 DMENESDKYFYFETPKACKACKINLIDDIKYFDE-----KGIMRTGLISFENNYYFNEN 2269
QY 1449 GQOAFNKSVTNKGTYVFGNDGTAQTA--GNPKGTFFDGSDIRFYFMSGQVLTGSGWYE 1506
Db 2270 GIIQFG-YINIEDKMFYFEGEDGMQIGVNTPDGPKYFAHQNTLDENFEGESINYTGW-- 2326
QY 1507 NAQGOWLYYKNGKVIITGLQTVGSRVYFDENGIIQAKGAVRTSDGKIRYFDENSGSMITN 1566
Db 2327 -----LGLD---EKRYFTDEVIATGTVI--IDGBEYFDPDPTAQVLIS 2366

QY 1567 Q 1567
Db 2367 E 2367

RESULT 4
US-11-013-759-11
; Sequence 11, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN OF MORAXELLA
; FILE REFERENCE: 1038-921MIS:jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 2314
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-11

Query Match 3.4%; Score 285.5; DB 7; Length 2314;
Best Local Similarity 19.0%; Pred. No. 1.2e-07;
Matches 340; Conservative 203; Mismatches 673; Indels 571; Gaps 77;

QY 29 SALGASVASADTTASD-----DSNQAVVTGDTTNNQATDQTSIAATATSEQASSTD 81
Db 792 NATTAKVYDETNQTSKVTVVNVDEKTIETLGDNGKTKIGVKTTLTTTNNAGKATNF 851
QY 82 AATDQ-----ASAEOTOGTTASTDQAAT-----TT 108
Db 852 STTDNDALVNAKDIAENLNTLAKEIHTKTGTADTALQTFVKKDGADTDETTVGGKGTQ 911
QY 109 NANEAKWVPTEENOGFTDEMLAEAKNVATAESDSIPSLAKMNVKQVDGKYYVDQDQ 168
Db 912 NGKTVNTLKLKGN-GLT-----VATNKDGTVTFGINTQSLKAGDST--TLNKDG 959
QY 169 ----NVKKNPAVSVG-DKIYFDETGAYKDTSKVDADKSSAVSQNATIPAAANNRAYSTS 223
Db 960 LSIKNPASNEQIQVAGGVKF-----AKVDKGNSTGIDGTSRI--TKDQIGFTGA 1008
QY 224 AKNFEADVNTLTADSWYRPSKILKDKTWTESGKDDFRPLLMAWPDTETKRNVNY-MN 282
Db 1009 NGSLDTTKPHLTOKD-----LKVG-----EVEITNTGINAGGK 1041
QY 283 KVVGIDKTYTAETSQADLTAABEL-VQARIEOKITSNTNKTWLRREALSAFVKTPQWN-- 339
Db 1042 KITNIQSGDITQNSNDAVGTGGRVYDLKTELESKINSA--AKTAQNSLHEFSVADEQGNHF 1099
QY 340 --GESEKPYDHLQNGALLPDNODLTPDQSNRYLLNRPTTQTSGLDS-RTYTPNDP 396
Db 1100 TVSNPYSYDTSKTSVITFAGENGIT--TKVNGGVV-RVGIDQTKGLTTPKLTGVNNG 1156
QY 397 LGGYDFLLANDVNSNPVQAEQLNMLHLLNFGSIY-----ANAD----- 438
Db 1157 KG----IVIDSQKQNTITGLS--NTLANVTNDGAGHALSQGLANDTKTRAASIGDVLN 1210
QY 439 -----ANPDSI---RVDADVNDADLLQISSDY-----LKAAYCIDKNNKNNHVS 482
Db 1211 AGFNLQNGEAVDFVSTYDVTVDIGNATTAKVTYDDTSTKTSKVYVDVNVDNKT----- 1264
QY 483 IVEAWSQ-----NDTPYLHDDGDLNLMNMDNKFRLSMLWSLAKPLDKRS 525
Db 1265 -IEVTSDDKLGKVTTLTKTSANGNATKFSAADG-----ALVKASDIAT 1308

QY 526 GLNPLHNSLVDRVDREVETVPSYSPARAHDSFVQDIIRDIKAEINPNSFGYSFTQE 585
Db 1309 HLNTLA-----GDIQTAKGASQASSASYV-----DADGNKVIYD----- 1343
QY 586 EIEQAFKIYNEDLKTKDKKYTHYNVPLSYTLLLTNKGSIPIRVVYGDMPFDDGQYMANKT 645
Db 1344 -----STDKKYYQVN-----DKQVDKNKEV 1364
QY 646 NYDAIESLLKARKMYVSGQAMONY-----QINGEILTSVRYKGKALKOSKGDAT 698
Db 1365 AKDK----LVAQAQTPDGTLAQMANVKSVINKEQVNDANKQOGINEDNAFIGLENAAXKT 1420
QY 699 RTSGVGVVMGN-----QPNFSLDCKVVALNMGAAHQAQYERALMVSTKDGVTATDA 751
Db 1421 KTKNAAVTVGDLNANAQVTELTTFAGDTGTTAKKLG-----TLTKGGQDTTNKL 1469
QY 752 DASKAGLVKRTDENGLYFLNDDL-----KGVANPOVSGFLOVWVPVGA 796
Db 1470 TNNIGVAGTD--GETVKLAKDLTNLNSVNAGGTIDEKGISFVDANGQAKANTPVL 1527
QY 797 DDQDI---RVAASDTASTDGKSLHQDAAMDSTRMFEFGFSNFQSFATKEBEYTNVIVANNV 853
Db 1528 NGLDLGGKRI SNIGAAVDDNDV-----NFKQFNEVAKTVNLANQNSNS 1571
QY 854 DKFVSWGITDFEMAPQYVSSDQQLDSV--IQNGYAFTRDYDLGMSKANKYGT-ADQLV 910
Db 1572 GASLPFVVTDANGK--INGTDGPKQKAIKAGADGKYHANANGVPVDKQKPIITDADKL- 1628
QY 911 KAIKALHAK---GLKVMADWVPDQMYTPPKQEVTVTR-----TDKFGKPIAG-- 955
Db 1629 -ANLAHGPPLDAGHQVVAS-----LGNSDAITLTWIKSTLPQIDPTNTGNANAGQA 1680
QY 956 -----SQINHSLYVTDTKSSGDDYQAKYG-----GAFIDELKEKY 990
Db 1681 QSLPSLSAAQQAASVKDLNVGNFLQTNHNVQDFVKAYDTVNFVNGTGADITSVRSAD 1740
QY 991 PELFTYKQISTQAI-----DPSVKIKQMSAKYF-----NGSNILGRGA----- 1029
Db 1741 G---TMSNITVNTALAAATDDGCVLIIKAKDGKPYKADDLMPNGSLKAGKSASAKTPTGL 1797
QY 1030 -----DYVLSQVSNKYF-----NVASDTLFLPSSLLGK-----VVE 1061
Db 1798 SLVNPAGKSGTGDAVALNNLSKAVPKSKDGTITTTVSSDGI-----SIQKONSSITLSK 1853
QY 1062 SGIRYDGK--GYIYNSSATGDQVKASFITEAGNLYYFGK-----DGYVMTGATNGAN 1113
Db 1854 DGLNVGKGVISNVGKGTQDAAANVQQLNEVRNLLGLGNAGNDNADGNQVNIADIKDPN 1913
QY 1114 YFFLENGTALRNTIYTDAGNSHYANDCKRYENGYYQFQGNQWRYFKDGNMAGVLTIV-- 1171
Db 1914 SGSSNRVTIKAGTVLGGKN-----NDTEKLATGQVQGVVD-----KGNANGDLNVMV 1964
QY 1172 -----DGNVOYFDKDGVOAKDKIIVTRDQKVRYPDQNGNAATNTFIA 1214
Db 1965 KTKQKSKALLATYNAAGQTYLTNPAEADRI---NEQGIRFFHVNDGNEP----- 2016
QY 1215 DKTGHYYLKGDAVAVTGAQTVGKQKLYFEANGQQVKGDPVTSDEGLKLYFYVDSDGDMWT 1274
Db 2017 -----VWQGRNGI---DSSASGKHSV---AIGFOAKAD-----GEAAV 2048
QY 1275 DTFIEDKAGWVFLGKDAVAVTGAQTVGKQKLYFEANGQQVKGDIYKGTGDKIRYDYDAKS 1334
Db 2049 AIGROTQAGNQSIAGIDNAQATGDQSI-----AIG---TGNVVTG-----KHS 2088
QY 1335 GEQFVNKTAKADGKTYVIGNDGVAVDPSVVKQTFPKDASGALRFYNLKGQLVTGSGWYE 1394
Db 2089 GAIGDPTVKA--DNSYSYVGNNOQFIDAT-----QTDVFGVNNITVTSNSVALGNSAI 2142
QY 1395 TA-NHDMVVIQSGKALTGQTINGQHLKPKB-DGHQVKGQLVGTGDTG-KVRYDYDANSQDQ 1451
Db 2143 SAGTHAGTQAKKSDGTAGTGTGTTTAGATGTGKGFAGTAVGAVSVGASGAEERRIQNVAAGEV 2202

QY 1452 AFNKSUTVNGKTYFNGDGTAGTGNPKGQTFKDGSDIRFYSMEGQ----- 1497
Db 2203 SATSTDAVNGSQLYKATQGIATNBE-----LDHRIHQENKANAGISSAMAWASM 2253
QY 1498 -----LVTGSGWYENAQ-----GOWLYVNGKVLIT 1522
Db 2254 PQAYIFGRSMVTGGIATHNGQGAVALGSLKSLSDNGQWVFKINGSADT 2300

RESULT 5
US-11-013-759-9
; Sequence 9, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; FILE REFERENCE: 1038-921MIS:jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; PRIOR FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 9
; LENGTH: 2053
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-9

Query Match 3.4%; Score 284.5; DB 7; Length 2053;
Best Local Similarity 19.8%; Pred. No. 1.2e-07;
Matches 348; Conservative 215; Mismatches 651; Indels 541; Gaps 87;

QY 31 LGASVASADTDITAS-----DSSNQAVVTGDTTNNQATDQTSIAATATSEQSASTDA 82
Db 336 INVAGSRDTDVAVQAQLKVLIELANRKITFKGDGDNNSVERGLNLTITIKGDAQTNA 395
QY 83 ATD-----QASABQTOGTASTDTTAAQTTHANAKWVPTENQGTDEML 130
Db 396 LFEANIGVVTGNGLVKLAKELTGLTSATNKITVSTNTNNN---ABLSGGLTFSP 452
QY 131 ABANWATAESDIPSDLAKMSNVKOVGGYKYDDODGNVKNFAVSGDKIYY----- 184
Db 453 TGTKTOKT-----VYSIDGLKFTNDSNSIATKGTTRITKKKIGFAGTNDG 497
QY 185 FDETGAUKDTSKVDADKS---SSAVSQNATI-----FAANNRAVSTSAKN----- 226
Db 498 VDESKPYLDNEKLKVGNSLNGSLTVNNTGNKQIQVANGIKFATVANNVANTSATVG 557
QY 227 -----FEAVDNYLTADSWYRPKSLKOGKT--WTESKDDPRLLMAWPDTEPK 274
Db 558 TARITEEKIGFAGTNDGVEQAPYLDKERLKVGRVEITTDG----- 599
QY 275 RNYAVNMN-KVVGIDKTYTAETSQADLTAAELVQARIIEQKITSNTKWLREASAPVK 333
Db 600 -----INAGNHKITGL-----TNGIANTDAVTIKQLKDAK----- 629
QY 334 TOPQWGESEKPYDDHLQNGALLFONQOTLTPTDQSNYRLLNRTPTNQTSLSRFTYNP 393
Db 630 -----PTLNA-----GDGISNSNNGDLV-DSSGNI-----TPTYNISVTKTL--NS 670
QY 394 NPLGGYDFLLANDVNSNPVQAEOLNWLHYLLNFGSIYANDADANFDSIRVADVNDV 453
Db 671 NCTSGNNKPSVSNADHNSLSLTAOLADYLNKV-----NETADSAFPSFKVQNGDSEN 723
QY 454 ADLLQISSD-----YLKAAYGIDKKNKNNHVSIV-VEAWSNDNTPYL----- 495
Db 724 -NAITVGKDTNGKTFNTLKLKGNGVNIITTNRATGTVTFGIDQSNGLATTPKLTGVS DNG 782

QY 496 -----HDDGDNLMNNDKFRLSMLWSLAKPLDKRSLNPLIHNLSLVREVDD--REV 545
Db 783 NRLVIRIQVPSADGNSTKNI-IKGLSPTLPSIASP-----SGRNIALGNTIIEEKQNAASI 837
QY 546 ETVPSSYSPARAHDSVQDIIRDIKAE-INPNSFGYSFTQEEIEQAFKI-YNEDLKKTKD 603
Db 838 DDVLNAGFNLKNGKDKDFVSTYDVFIDGNATTATVTYDEANQTSKVAYDYNVDEKTI 897
QY 604 KYTHYN-----VPLSYTLTLLTNKSGIPRVYVGDMDFTDDGQYMANKTVNYDAIESLLKA 656
Db 898 ELTGDMGKKQLGVKTIKLTETSTNGNA-----TTFSTDDHVALKASDIAGNLNLTAE 950
QY 657 RMKYVSG--GQAMQNYQIGNGEILTSRVYKKGALKQSDKGDATTR-TSGVGVVMGQPNF 713
Db 951 EIHHTKGTANTALQTFVKK-----VDENDKADDTNATITVGKDGTSKGVNTL 997
QY 714 SLDGKVVALMGAHANQBYRALMVST-KGVATYATDADAS-KAG-----LVKR 761
Db 998 KLKGGK-----NGLDIKTDKGTVTFGINTQSGIKAGDSTTLNNGLSIKN 1042
QY 762 TDENGYLEFLNDDLK-----GVANPOVSGFLQV-----WVPVGAADQDQIRVAASDTA 809
Db 1043 TASNEQIQVAGGVKPFAMVNGVVGAGIDGTTTRITRDEIGFTGTNGSLDKSKPHLSKDI 1102
QY 810 STDGKSLHQDAAMDGRVMEFSGFATKEEYTNVVIANNVDKFSVSGITDFEM-AP 868
Db 1103 NAGGKKI-----TNIQS-----GEIAKNSHDAVTGKGIYDLKTELE 1138
QY 869 QVVSST-----DQOFLDSVIONGYAFTDRYDLGMSK-----ANKYGTADQL 909
Db 1139 NKISSTAKTAQNSLHEFSVADEQGNFTVSNPYS---SYDTSKTSVDITFAGENGITTKV 1195
QY 910 VKAIKAL---HAKGLKVMADWVPDQWYTPPKQEVWTVTRTDKFGKPIAGSIQNHSLYVTD 966
Db 1196 NKGVRVIGIDQTKGL-----TTPK-----LTGVNNGKGIIVINSQNGQWITIG 1238
QY 967 -----TKSSGDDYQAKYGAGFLDELKERYPELFTKKQI-----STGQAIDPSVKIKQ 1013
Db 1239 LSNLTANVTNDKGSVRTTEQGNIIKDEDKTRAASIVDVLASAGFNLGNGEAVD---FVST 1295
QY 1014 WSAKIFNGSNILGRADYVLSQVSNKYF--NVASDTLFLPSSLKGVESGIRYDGKG- 1070
Db 1296 YDTVNFANGNTTAKVTYDDTSKTSKVVDVNVDDTTIEVKDKLKG-VKTTTLTSTGTGA 1354
QY 1071 --YIYNSAATDQ-KVASPI-----TEAGNLVYFGKDGVMVTGAQTINGANYFFLENGTA 1122
Db 1355 NKFSALNQATGALVKSADIVAHLNLTSGDI-----QTAKAS-----QA 1394
QY 1123 LENTTIYTDAGNSHYANDKREYNGYQOFGNDWRYFKDGNMAVGLTTVDGNVQYFDDKG 1182
Db 1395 NNSAGVVDADGNKVIYDST---DNKYQAKN-----DGTV---DKTK 1430
QY 1183 VOAKDKIIV---TRGKVRYFDQHNGNAATNTFIADKTGHYYLKGDKGVAVTG---AQTV 1236
Db 1431 EVAKDKLVAQAOTPDGTLA---QMNVKSVINKEQVNDANKKQGINEDNAFVAGLEKAASD 1487
QY 1237 GKQKLYFEANGQOVKQDFVTSDEGKLYFVDVDSGDWMTDTFIEDKAGNFWYLGKDGAAVT 1296
Db 1488 NKTK-----NAAVTVGDLNVAQAOTPLTF-----AGDTGTGA---KGLG-----EVLTK 1528
QY 1297 GAQTIIRGQKLYFKANGQOVKQDGVKGTGDKIRYDAKSGEQVFNKTVKAADGKTVY--IG 1354
Db 1529 GQQTDTN-----KLTNNI--GVVAGTDG-----FTVKLAKDLTNLSVN 1566
QY 1355 NDGVAVDPSVVKQTFKQASGALRF-----YNLKGQVLT--GSGHYETANHDDWYI 1403
Db 1567 AGGTKIDE---KGISFVDANGQAKANTPVSANGLDLGKVISNVGKGTGKTDAAANVQOL 1623
QY 1404 QSGKALTGEQTINGOHLVFKEDGHQVKGQLVGTGDKVRYDANGSDQAFNKSIV-----T 1458
Db 1624 NEVRNLLGLGNDNA-----DGNQ-----VNIADIK---KDPNSSGSS--NRTVIKAGTV 1667
QY 1459 VNGKTYFNGDGTAGTA-----GNPKG-----QTFKDGSDIRFYSMEQVLT 1500

Db 1668 LGK-----GNDETEKATGQVGVVDKGNANGDLNWNVVKTKQG-----SKALLAT 1717
QY 1501 GSGWYENAOQMLYVRNGKVLGTGLQTVGSRVYFDENGIGQAKGKAV-RTSDGKIRYFDEN 1559
Db 1718 -----YNAAGQTYVNN-----PAAEDRINEQGIREFHVN 1749
QY 1560 SGMITNQNWKFFVQO 1574
Db 1750 DG-----NOBPFVQGR 1760
RESULT 6
US-11-013-759-3
; Sequence 3, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN OF MORAXELLA
; FILE REFERENCE: 1038-921MIS:jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentJn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1992
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-3
Query Match 3.3%; Score 279.5; DB 7; Length 1992;
Best Local Similarity 19.5%; Pred. No. 2.1e-07;
Matches 354; Conservative 232; Mismatches 648; Indels 579; Gaps 91;
QY 23 SATMLASALCA-----SVASADTDTASDSDNQAVVTGD-----OTTNNQATD 64
Db 205 SSSGLAIALGAGTGAQLOQSIALGQGSVVTQSDNNSRPAYPTNQALDPKFAQTNNTKAG 264
QY 65 QTSIAATATSEQASATDAATDQASA-----ABEQGTGTASTDTAAQTNNANE 112
Db 265 PLSIGNSIKRKLIINVAGVNTKDAVNVAQLEAVVWKAKERRITFGQDDNSTDVKIGLDN 324
QY 113 AKWVPTENEOGFTDEMLAEAKNVATAESDIPSDLAK-MSNVKQYDQKYYVYDQGNVK 171
Db 325 TLTIKGGAETNALTDNNGVVK-----ADNSGLKVKLAKTLNLTNLTN-----TTTLN 373
QY 172 KNFAVSVGDKIYVFDGTGAVKDTSKYDADKSSSAVSONATIFAANNRAYSTSAKNFEAVD 231
Db 374 ATTTVKVGS-----SSTTAELLSLSLFTQNTQSGTSKTVGVGVNGKVTNNARETTAAIG 430
QY 232 NYLTADSWYRPSILKDGKWTBESGKDFRPLLMAMWPDTEKRNVTNNYNNKVVG-----286
Db 431 T-----TRITRDKIGFARDGVD-----EKQAPYLDKQLKVGSAIT 468
QY 287 IDKTYTAETSQADLTAAAEVLQARIEQKITSENNTKWLREASISAFVKTQPOWNGESEKPY 346
Db 469 IDNGIDAGNKKISNLA-----KGSSANDAVTIEQLKAA-----KP- 503
QY 347 DDHLQALLFDNQDOLTDPDTQS-----NYRL-LNKTPTNQTSGLDSRFTYNNPDPLGG 399
Db 504 --TLNAGAGISVPTSEISVDKSGNTAPTNYINGVKTTELSNDSGTS-D-KFSVKGS-----555
QY 400 YDFLLANDVDSNPVVOAEOQLNHLNFGSIYAND-----ADANFDSIRVDVNDVAD 455
Db 556 -----GTNNSLVTAEHL-----ASYLNEVNRNADTADALSQSFVKEEDDDAN 596
QY 456 LLIQSSDYLKAA-----YGIDK-----NN-----474

Db 597 AITVAKDITKARGAVSILKLGKNGLTVAATKKDGTTFGLSQDSGLTICKSLINDGLTV 656
QY 475 KNANHVSI-----VBAWSNDPTPYLHDDGD 500
Db 657 KDTNEQIQVGANGIKFTNVNGSNPGTANTARITRDKIGFAGSDGAVDTNKPYL--DQD 714
QY 501 NLMMNDKFERLSMLWSLAKPLDKRSLNPLHNSLVDREVDDREVE---TVPSYSFARAH 557
Db 715 KLOVGNVKITNTGINAGGKAI---TGLSPTL-PSIADQ---SSRNELGNTIOXKD--KSN 766
QY 558 DSEVOIIRDIKABINPNSPGYSFTOBEIEQA-----FKIYNEDLKKTDKKYTHYNVP 611
Db 767 AASINDILNTGFNLKNNNPIDFVSTYDIVDFPANGNATTATVTHDTANKTSKVVDVNVVD 826
QY 612 LSYTLLLTNKGSIPIRVYVYDMFTDDQYMWANKTVNYDAIESLLKARMKVYSGQGMQNYO 671
Db 827 -DTTIHLTG-----TDDNKKLGKVTTKLN-----KTSANGNTATNFN 862
QY 672 IGNG--EILTSVRYGKGALK-----QSDKGDATTTSVGVVGMGNQPNFSLDGKVVALN 723
Db 863 VNSDEDEDALVNAKDIAENLNTLAKIEHTTKGTADTALQTFVTKKVDENNADDAN--AIT 920
QY 724 MGAHAHNOEYRALMVSTKDG VATYATDADASKAGLVKRTDENGILYF-----770
Db 921 VGQKANNQVNTLTUKGN-----GLNIKTDKNGTVTFGINTTSLGKAGKS 966
QY 771 -LNDLLKGVANPOVSGFLQVWVP-----VGAADDQDIRVAASDT--ASTDGKS 815
Db 967 TLNDGGLSIKNPTGSEQIQVGADGVKFAKVNNGVVGAGIDGTTRITRDEIGFTGNG-S 1025
QY 816 LHQDAAMDSRVMFEGFSNFQSPATKEEYTNV-----VTANNVDFKFSWGITDEM-APQY 870
Db 1026 LDKSKPHLSK---DGIN-----AGGKITNTIQSGEIAQNSHDAVTVGGKIDYLDKTELENK 1076
QY 871 VBST-----DQGLDSVIONGVAFTRDYLGMSEK-----ANKYGTADQLVK 911
Db 1077 ISSTAKTAQNSLHESVADEQGNNTVSNPYS---SYDTSKTSVITFAGENGITTKVVK 1133
QY 912 AIKAL---HAKGLKVMADWVPDQMYTFPKQEVVTVTRTDKFKPIAGSQINSHLYVTD-- 966
Db 1134 GVVRVGIDQTKGL-----TTPK-----LTGVNNGKGIVIDSQNGQNTITGLS 1176
QY 967 -----TKSGDDDYQAKYGGAFDELKEYPELFTKKQI-----STGQAIDSVKIKQWS 1015
Db 1177 NTLANVTNDKGSVRTTEQGNIIKDEDKTRAASIVDVLSSAGFNLQNGGEAVD---FVSTYD 1233
QY 1016 KYFNGSNILGRGADYVLSDOVSNKYF--NVASDTLFLPSSLGKLVESGIRYDGKG--- 1070
Db 1234 TVNFADGNATTAKVYDDTSKTSKVVDVNVDDTIEVKDKKLK-VKTTTLTSTGTGANK 1292
QY 1071 TYNSSATGQ--VKASFI-----TEAGNLVYFGKDGVMVTGAQTINGANYFFLENTALR 1124
Db 1293 FALSNOATGDLVKASDI VAHLNTLSGDI-----QTAKGAS-----QANN 1332
QY 1125 NTIYIDAQNSHYANDKRYENGYYQOFGNDWRYPKDGNMAGVLITVDGNVQYFDKGQVQ 1184
Db 1333 SAGYVDADGNKVIYDST---DNKYQAKN-----DGTV---DKTKEV 1368
QY 1185 AKDKTIIV---TRDGKRVYFDQHNGNAATNTFTADKTGHWYILGKGVAVTG---AOTVKG 1238
Db 1369 AKDKLVAQAQTPDGLA---QNNVKSVINKEQVNDANKQGINEDNAFVKGLEKASDNK 1425
QY 1239 OKLYFEANGQQVKGDFVTSDEGLKLYFYDVDSGDMTDTFIEDKAGNFWLKGDKGAATGA 1298
Db 1426 TK-----NAAVTVGDLNAVAQTPLTF---AGDTGTTA---KKLG-----ETLTIKGG 1466
QY 1299 QTIROGLKLYFKANGQQVKDGIUKGTGKIRYDADKSGEQVFNKTKAADGKTVV--IGND 1356
Db 1467 QTDITN-----KLTDNNI--GVVAGTDG-----FTVKLAKDLTNLNSVNAQ 1504
QY 1357 GVAVDPSVVKGOTFKDASGALRF-----YNLKGQLVT--GSGWYETANHWDVVIQIS 1405

Db 1505 GTKIDD---KGVSFVDSGQAKANTPVLSANGLDLGKVI SNVCKGKYOTDAAANVQOLNE 1561
QY 1406 GKALTGBQTINGOHLYPEGQHVKGQLVTGTDGKVRYYDANSQAFNKSVD---TVN 1460
Db 1562 VRNLLGLGNAGNDN---ADGNQ---VNIADIK---KDPNSGSSS-NRTVIKAGTVLG 1608
QY 1461 GKTYFYGNDCTAQT---CNPKG-----QTFKGSDFRFSMEGOLVTGS 1502
Db 1609 GK---GNNDTEKLATGGIOGVVDKQGNANGDLSNVWVKTKQDG-----SKALLAT-- 1656
QY 1503 GWYENAGQWLYYVKNKGKVLTLGTLQVRSQVYFDENGIOAKGKAV-RTSDGKIRYFDENSG 1561
Db 1657 ---YNAAGQTNLYTNN-----PAAIDRINEQIRFFHVNDG 1690
QY 1562 SMITNWKFFYQQ 1574
Db 1691 ---NOEPVVQGR 1699

RESULT 7
US-11-013-759-13
; Sequence 13, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Kleih, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN OF MORAXELLA
; FILE REFERENCE: 1038-921MIS:jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; PRIOR FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 1992
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-13

Query Match 3.3%; Score 279.5; DB 7; Length 1992;
Best Local Similarity 19.5%; Pred. No. 2.1e-07;
Matches 354; Conservative 232; Mismatches 648; Indels 579; Gaps 91;
QY 23 SATMLASALGA-----SVASADTDASDDSNQAVVTGD-----QTTNQAOTD 64
Db 205 SSSLGAILGAGTRAQLOGSIALGOGSVVVTQSDNNSRPAYTPNTQALDPKFOATNNTKAG 264
QY 65 QTSIAATATSEASATDAATDQASA-----ASQTQCTTASTDTAAQTNNANE 112
Db 265 PLISIGNSIKRKLIINVAGVKNKTDVNVVAQLEAVVVKAKERRITFGDDNDSTDKVIGLDN 324
QY 113 AKWVPTENENOGFTDEMLAEAKNVATAESDIPSLAK-MSNVKQVDGKYYYYDDGKNVK 171
Db 325 TLTIKGAETNALDNNIGVKE---ADNSGLKVKLAKTLNLTENV-----TTILN 373
QY 172 KNFAVVGDKIYYFDETGAYKQTSKVDADKSSAVSQNATIFPAANNRAYSTSAKQPEAVD 231
Db 374 ATTTVKVGSS---SSTTBALLSDSLTFTQPTNGSQSTSKTVGVNGVKFTNNAETAAIG 430
QY 232 NYLTADSWRPKSIKDKGTWTESGKDDFRPLLMAWWPDTETKENVYNNVKNVVG----- 286
Db 431 T-----TRITRDKITGFARDGDVD-----EKOAPYLDKQKLVGSGVAIT 468
QY 287 IDKTYTAETSQADLTAAAEVLQARIQKITSENNTKWLREAIASAFVKTPQWNGESEKPY 346
Db 469 INDGIDAGNKKIISNLA-----KGSSANDAVTIEQLKAA-----KP- 503
QY 347 DDLHQALLFDNQDLDLTPDTQS-----NYRL-LNRFTPTNQTSGLDSRFTYNNPDPLGG 399

Db 504 --TLNAGAGISVTPTEISVDAKSGNVTAPTNYIGVKTTELNSDGTSD-KFSVKS----- 555
QY 400 YDFLLANDVDSNPVQAEOLNMLHYLLNFGSIYAND---ADANPDSIRVDAVDNVDAD 455
Db 556 -----GTNNSLVTAEHL-----ASYLNEVNRATASALQSFTVKEEDDDAN 596
QY 456 LLOISSDYLKAA-----YGIDK-----NN----- 474
Db 597 AITVAKDITTKNAGAVSILKLGKNGLTVAATKDGTVTFGLSQDSGLTIGKSTLNNDLGLTV 656
QY 475 KNANNHVSI-----VEAWSNDNTPYLHDDGD 500
Db 657 KOTNEIOIGVANGIKFTNVNGSNPGTGIANTARITRDKIGFAGSDGAVDTNKPYL--DQD 714
QY 501 NLNMNDNKFRLSMLSLAKPLDKRSLNPLIHNLSVDREVDREVE---TVPSVSAPARAH 557
Db 715 KLOVGNVKITNTGINAGGKAI---TGLSPTL-PSIADQ---SSRIELGNTIQKD--KSN 766
QY 558 DSEVDQDIIRDIKAEINPNSFGYSFTQEEIEQA-----FKIYNEDLKTKDKYTHYNVP 611
Db 767 AASINDILNTGPNLKNNNPIDFVSTYDIDVDFANGNATTATVTHDTANKTSKVYVDVNDV 826
QY 612 LSYTLLLTNKGSIPTVYVYGDMTDDQOYMANKTVDYATIESLLKARKVYVSGQOAMQNYQ 671
Db 827 -DTTHLTG-----TDDNKLGVKTKTLN-----KTSANGNTATNFN 862
QY 672 IGNG--EILTSVRYGKALK-----QSDKGDATTRTSGVGVVMGNQPNFSLDGKVVVALN 723
Db 863 VNSDDEDALVNAKDIAENLNTLAKEIHTTKGTADTALQTFTVKKVDENNADDAN--AIT 920
QY 724 MGAHAHOEYRALMVSTKDGVAITYATADAKASKAGLVKRTDENGILYF----- 770
Db 921 VGQKNANNOVNTLTGKEN-----GLNIKTDKNGTVFINGITTSGLKAGKS 966
QY 771 -LNDLLKGVANQVQSFLOVWVP-----VGAADDQDIRVAASDT--ASTDGKS 815
Db 967 TLNDGGLSINKPTGSEQIQVGADGVKFAKVNNGVVGAGIDGTTRITRDEIGTGTNG-S 1025
QY 816 LHQDAAMD SRVMFEGFSNFSPATKEEYTNV---VIANNVDKFSWGTIDDEM-APQY 870
Db 1026 LOKSPHLSK---DGIN-----AGKKITNQSGEIAQNSHDAVTGKGKIYDILKTELENK 1076
QY 871 VSST-----DQFLDSVIONGYAFTDRYDLGMSK-----ANKYGTADQLVK 911
Db 1077 ISSAKTAQNSLHEFSVADEQGNFTVSNPYS---SYDTSKTSVDITFAGENGITTKVKN 1133
QY 912 AIKAL--HAKGLKMWADWVPQMTYTFPKQEVVYVTRTDKFGKPIAGSQINHSLYYVD-- 966
Db 1134 GVVRVGIDQTKGL-----TTPK-----LTVGNNGKGIVIDSQNGQNTITGLS 1176
QY 967 -----TKSSGDDYQAKYGGAFDELKEKYPELFTKKQI-----STGQALDPSVKIKQWS 1015
Db 1177 NTLANVTNDKGSVRTTEQGNIIKDEKTRAASIVDLSAGPNLQGNGEAVD---FVSTYD 1233
QY 1016 AKYFNGSNILGRGADYVLSQVSNKYF--NVASDITLFLPSSLGKVVESGIRYDGGK--- 1070
Db 1234 TVNFADGNATTAKYVYDDTSTKTSKVYVDVNDVDDTIEVKKDLG-KVTTLTISTGTGANK 1292
QY 1071 YIYNSSATGQD-VKASFI-----TEAGNLYYFGKDGWYVTGAQTINGANYFFLENGTALR 1124
Db 1293 FALSNOATGDALVKASDIAVAHLNTLSGDI-----QTAGKAS-----QANN 1332
QY 1125 NTIYTDAGNSHYVANDGKRYENGVOQFGNDWRYFKDGNMVAGLTIVDGNVQYFDKGVQ 1184
Db 1333 SAGYVDADGNKVYDST---DNKYQAKN-----DGTV---DKTREV 1368
QY 1185 AKDKIIV---TRDGKVRVFDQHNNAATNTFIADTKGHVYILGKGVAVTG---AOTVGK 1238
Db 1369 AKDKLVAQAQTPDGLA---QNNVKSVINKEQVNDANKQGINEDNAFVKGLEKASDNK 1425
QY 1239 OKLYFEANGQVKGDFVTSDCKLYFYDVDSDGMWTDTFIEDKAGNFWYLGKGAAVTGA 1298
Db 1426 TK-----NAAVTVGDLNVAQTPLTf-----AGDTGTGA---KKLG-----ETLTKG 1466

```
QY 1299 QTRGOKLFPKANGQOVKGDIIVKGTGKIRYYDAKSQGEQVFNKTVKAAADKTVV--IGND 1356
Db 1467 QTDTN-----KLTDDNI--GWAGTDG-----FTVKLAKDLTNLSNVAG 1504
QY 1357 GVAVDPSVVKGOFPKASGALRE-----YNLKGQLV--GSGWYETANHHDWYVYIQS 1405
Db 1505 GTKIDD---KGVSVFVSSGQAKANTPVLNANGLDLGGKVISNVGKTKDTDAANVOQLNE 1561
QY 1406 GKALTGEOTINGOHLYFKEDGHQVKGQLVTGTDGKVRYYDANGSDQAFNKSIV-----TVN 1460
Db 1562 VRNLLGLGNAGNDN---ADGNQ-----VNIADIK--KDPNSGSSS-NRTVIKAGTVLG 1608
QY 1461 GKTYFYGNDGTAGTA-----GNPKG-----QTFKDGSDIRFYSMEGQLVTS 1502
Db 1609 GK-----GNNDTEKLTGCGIYGVDKDNANGDLSNVWVKTKDQ-----SKALLAT-- 1656
QY 1503 GWYENAGQOWLYVYKNGKVLTLGLTVGVSQRVYPBENGIOAKGKAV--RTSDGKIRYFDNSG 1561
Db 1657 ---YNAAGQTYLUTN-----PAEIDRINEQGRFFHVNDG 1690
QY 1562 SMITNQMKFYVG 1574
Db 1691 ---NQEPVQGR 1699

RESULT 8
US-11-013-759-4
; Sequence 4, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN OF MORAXELLA
; FILE REFERENCE: 1038-921MIS:jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2047
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-4 )

Query Match 3.3%; Score 279.5; DB 7; Length 2047;
Best Local Similarity 19.5%; Pred. No. 2.2e-07;
Matches 354; Conservative 232; Mismatches 648; Indels 579; Gaps 91;

QY 23 SATMLASALCA-----SVASADTDTSDDSNQAVVTD-----QTTNQATD 64
Db 260 SSSLGAIAGLAGTRAQLQGSALVQGSVVTSQSDNNRSPAYTPNTQALDPKFAATNNKAG 319
QY 65 QTSIAATATSEQSASTDAATDOASA-----AEQTQGTASTDTAAQTNNANE 112
Db 320 PLSIGNSIKRKLINVGAGVKNKTDVNVQAQLEAVVWAKERRITFOGDDNSTDVKIGLDN 379
QY 113 AKWVPTENQGTDEMLAEAKNVAEASDPSIDLAK-MSNVQVQDGKYYVYDQGNVK 171
Db 380 TLTIKGAETNALTDDNIGVVKE---ADNSGLVKVLAETLNNLTENV-----TTTLN 428
QY 172 KNFAVSGDKIYFDETGAKYKOTSKVDADKSSSAVSQNTAIPANNRAYSTSAKNEAVD 231
Db 429 ATTVKVGSS---SSSTAELLSLTFPTQNTQSGTSKTVYGVNGVKFTNNAETAAIG 485
QY 232 NYLTADSWVRPKGILKDGKTWESGKDDFRPLLMAWMPDPTETKRNYYVNMKNVVG----- 286
Db 486 T-----TTRTDKIGFARDGVD-----EKQAPYLDKKQKLVGSVAIT 523
```

```
QY 287 IDKTYTAETSOADLTAAAEELVOARIEOKITSENNTKWLBEAISAFAVKTOPOWNGESEXKY 346
Db 524 IDNGIDAGNKKITSNLA-----KGSSANDAVTIEQLKAA-----KP-- 558
QY 347 DDHLQNGALLFONQOTDLTPDQTOS-----NYRL-LNRTPNTQOTSLDSRFTYPNPNDPLGG 399
Db 559 ---TLNAGAGISVTPTEISVDAKSGNVNPTVYNIKVKTTELNSDGTSD-KFSVKGS----- 610
QY 400 YDFLLANDVNSNPPVQABQLNWLHLYLNFSGSIYAND-----ADANPDSIRVDAVNDVAD 455
Db 611 -----GTNNSLVTAEHL-----ASYLNEVNRATDSALQSFTVKEEDDDAN 651
QY 456 LLQISSDYLKAA-----YCIDK-----NN----- 474
Db 652 AITVAQDITTKNAGAVSILKKGKNGLTVATKKDGTVTFGLSQDSGLTIGKSLTNGLTV 711
QY 475 KNANNHVSI-----VEAWSNDPTPYLHDDGD 500
Db 712 KDTNEQIQVGANGIKFTNVGNSNPGTGIANTRDKIGFAGSDGAVDTNKPYL--DQD 769
QY 501 NLMMNDNKERLSMLMSLAKPLDKRSGLNPLIHNSLVDRVDDREVE---TVPSYSFARAH 557
Db 770 KLVQGNVKITNTGINAGGKAI--TGLSPTL-PSIADQ--SSRNIELGNTIQDKD--KSN 821
QY 558 DSEVQDIIRDIKAEINPNSFGYSTQBEIEQA-----FKIYNEDLKKTKDKKYTHYVNP 611
Db 822 AASINDILNTGPNLKNNNPIDFVSTYDIVDFANGNATTATVTHDTFANKTKVYVDNVND 881
QY 612 LSVYTLNLTNKGSIPIVYVYGDMTDDGOYMANKTVNYDAIESLLKARMKVSVSGQAMQNYQ 671
Db 882 -DTHIHLTG-----TDDNKKLGCVTKTLN-----KTSANGNTATNFN 917
QY 672 IGNG--EILTSVRYGKGALK-----QSDKGDATRTTSGVVMGMGNPFLSDGKVVALN 723
Db 918 VNSDEEDALVNAKDAENLNTLAKEIHTTKGTADTALQTFTVKVKVDENNADDAN--AIT 975
QY 724 MGAHANQBYRALMVSTKDGVAITYATDADAKAGLVKRTDENGILYF----- 770
Db 976 VGQKNANNQVNTLTLLKGEN-----GLNIKTKNGTVTFGTINTTSLGKAGKS 1021
QY 771 -LNDDLKGVANPQVSGFLQVWPV-----VGAADDQDIRVAASDT--ASTDGKS 815
Db 1022 TLNDGGLSINKNPTGSEIQIUGADGVKFAKVNNGVVGAGIDGTTRITREIEGFTGNG-S 1080
QY 816 LHQDAAMDSTRVMPFEGSFNSQSPATKEEYTNV-----VIANNVDKFKVSWGITDEM-APQY 870
Db 1081 LDKSKPHLSK--DGIN-----AGGKITNIQSGEIAQNSHDAVTVGGKIYDLKTELENK 1131
QY 871 VSST-----DGOFLDSVIQNGYAFTRDYDLGMSK-----ANKYGTADQLVK 911
Db 1132 ISSHTAKTAQNSLHEFSVADEQGNNTFVSNPYS---SYDTSKTSDDVTTFAGENGITTKVKN 1188
QY 912 AIKAL---HAKGLKVMADWVPQMVTFPKOEVVVTVTRTDKFGKPIAGSQINHSLYVTD-- 966
Db 1189 GVVRVGIIDTKGL-----TTPK-----LTVGNNGKGIVIDSQNGQNTITGLS 1231
QY 967 -----TKSSGDDYQAKYGGAFDELKEKYPFLTKKQI-----STGQAIIDPSVKIKQWS 1015
Db 1232 NTLANVTNDKGSVRTTEQGNIIKDEKTRAASIVDVLSAGFNLOQNGEAVD---FVSTYD 1288
QY 1016 AKYFNGSNLGRGADVVLSDQVSNKYF--NVASDRTLFLPSSLLGKVVESGIRYDGGK--- 1070
Db 1289 TVNFADGNATTAKTVYDDTSKTSKVYVDVNVDDTTIEVKDKKLG-VKTTTLTSTGTGANK 1347
QY 1071 TYVNSSATDQ-QKASFI-----TEAGNLYYFGKQGYVMTVGAQTINGANYFFLENGTALR 1124
Db 1348 FALSNOATGDALVKASDIIVAHNLTLSGDI-----QTAGAS-----QANN 1387
QY 1125 NTIYTDAGNSHYANDGKRYENGVOQFGNDMRYFKDGNNAVGLTTVDGNNVOYFDKDGQV 1184
Db 1388 SAGYVDADGNKVIYDST---DNKYYQAKN-----DGTV---DKTKEV 1423
```

QY 1185 AKDKIIV---TRDKVRYFDQHGNGNAATNTFIADKTHGWHYVLGKGVAVTG---AQTGK 1238
Db 1424 AKDKLVAQAQTPDGTTLA---QMNYSVINKEQVNDANKKQGINEDNAFVKGLEKAASDNK 1480
QY 1239 QKLYFEANGQOVKQDFVTSDEGKLYFYDVDSGDMWTDTTFIEDKAGNWFYLGKOGAAVTGA 1298
Db 1481 TK-----NAAVTGDLNVAQTPLTF---AGDTGTTA---KKLG-----EFLTINGG 1521
QY 1299 QTRGKLYFKANGQOVKQDIKGTGDKIRYDAKSGEQQVFNKTVKAADGKTYV--IGND 1356
Db 1522 QTDN-----KLTDNNI--GVVAGTDG-----FTVKLAKDLTNLSNVAG 1559
QY 1357 GVAVDPSPVKGOTFKDASGALRF-----YNLKGQLVT--GSGHYETANHDWVVIQS 1405
Db 1560 GTKIDD---XGVSFVDSGQAKANTPVLSSANGLDLGGKVISNVGKGTDKTDAANVQQLNE 1616
QY 1406 GKALTGQTINGOHLYFEKEDHGVKQGLVTGTGDKVRYDANSQDAFNKSV-----TVN 1460
Db 1617 VRNLLGLGNAGDN---ADGNQ-----VNIADIK---KDPNSGSSS-NRTVIKAGTVLG 1663
QY 1461 GKTYFFONDGTAQTA-----QNPKG-----QTFKQGSDIRFYSMESQQLVTGS 1502
Db 1664 GK-----GNNDTEKLATGCIQGVVDKOGNANGDLSNVVWKTKQD-----SKKALLAT-- 1711
QY 1503 GWYENACQWLYVNGKVLTLGLQTVGSQRYVFBENGIOAKGKAV-RTSDGKIRYFDNSG 1561
Db 1712 ---YNAAGQNYLTNN-----PAAIDRINEQGIREFHVNDG 1745
QY 1562 SMITNQWKFVYGQ 1574
Db 1746 ---NQEPVQGR 1754

RESULT 9

US-11-013-759-7
; Sequence 7, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loomore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; FILE REFERENCE: 1038-921MIS:jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 7
; LENGTH: 2047
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-7

Query Match 3.3%; Score 279.5; DB 7; Length 2047;
Best Local Similarity 19.5%; Pred. No. 2.2e-07;
Matches 354; Conservative 232; Mismatches 648; Indels 579; Gaps 91;

QY 23 SATMLASALGA-----SVASADTDPTASDSDSNQAVVTGD-----QTTNQAOTD 64
Db 260 SSSLGAILAGTAGTQAQLOSGIALGQGSVVVTQSDNSRNPATPTQALDPKFOATNTKAG 319
QY 65 QTSIAATATSEQASDTAANDQASA-----ABEQGTGTASTDTAAQTITNAME 112
Db 320 PLSIGSNSIKRKIIINGAGVKNKTDVAVNVAQLEAVVWAKERRITFGQDDNSTDVKIGLDN 379
QY 113 AKWVPTPENOGFTDEMAEAKVATAESDPSDLAK-MSNVQYDQKYYDDQGNVK 171
Db 380 TLTIKGAETNALTDNNIIGVWKE---ADNSGLKVLKLAKTLLNLTENV-----TTLLN 428

QY 172 KNFAVSVGDKIYYFDETGAYKDYTSKVDADKSSSAVSQNAITFAANNRAYSTSAKNFEAVD 231
Db 429 ATTIVKVGSS---SSTTAELLSDSLTFQPNVTGSGSTSKTVYGVNGVKFTNNAETTAIG 485
QY 232 NYLTADSWYRPKSLTKDGKWTESGKDDFRPLLMAMWPDTEKRYVNYNMKVVG-----286
Db 486 T-----TRITRDKIGFARDGDVD-----EKQAPYLDKQKQKLVGSAIT 523
QY 287 IDKTYTAETSQADLTAAAEVLQARIEQKITSENNTKWLREAIASAFVKTQPMNGESEKPY 346
Db 524 IDNGIDAGNKISNLA-----KGSANDAVTIEQLKAA-----KP- 558
QY 347 DDHLQALLFDNQDTLPTQOS-----NYRL-LNRPTNQTSGLSRLFRYPNPDPLGG 399
Db 559 --TLNAGAGISVTPTEISVDAKSGNVTAPTYNIGVKTTLSNDSQTS-D-KPSVKGS-----610
QY 400 YDFLLANDVNSNPVQAEOLNWLHLLNFGSIYAND-----ADANFOSIRVDVNDVAD 455
Db 611 -----GTNNSLVTAEHL-----ASYLNEVNRTADSALQSFTVYKEEDDDAN 651
QY 456 LLOISSDYLKAA-----YGDK-----NN-----474
Db 652 AITVAKDTTKNAGAVSILKLGKGLTVATKDGTVTFGLSQDSGLTIGKSTLNDGLTV 711
QY 475 KNANHHVSI-----VEAMSDNDTPYLHDDGD 500
Db 712 KDTNEQIQVGANGIKFTNVNGSNPGTGIANTRITRDKIGFAGSDGAVDINKPYL--DQD 769
QY 501 NLMNMNKFRLSMLWSLAKPLDKRGLNPLIHNSLVDREVDDREVE---TVPSVSPARAH 557
Db 770 KLGQVNVKINTNGINAGGKAI---TGLSPTL-PSIADQ---SSRNIELGNTIQDKD--KSN 821
QY 558 DSEVQDIIRDIKAEINPNSFGYSFTQEEIEQA-----FKIYNEDLKKTKKTYHNVP 611
Db 822 AASINDLNTGFLNKNNDPIDFVSTYDIVDFANGNATTATVTDTANKTSKVYDVNVVD 881
QY 612 LSYTLTLTKGSIPIRVYVYGMFTDDGQYMWANKTVNYDAIESLLKARKMYVSGGQAMQNYQ 671
Db 882 -DTHILTG-----TDDNKKLGKVTTKLN-----KTSANGNTATNPN 917
QY 672 IGNG--EILTSVRYGKALK-----QSKGDATRTSGVGVVMGNOPNLSLQKVVVALN 723
Db 918 VNSSDEDALVNAKDIAENLNTLAKEIHTTKGTADTALQTFVTKKVDENNADDAN--AIT 975
QY 724 MGAHAHQEYRALMVSTKGVATYATDADASKAGLVKRTDENGVLVYF-----770
Db 976 VGOKNANNQNTLTLKEN-----GLNKTDKGTGTVTFGINTTSGLKAGKS 1021
QY 771 -LNDDLKGVANQVSGFLQVWVP-----VGAADDDQDIRVAASDT--ASTDGKS 815
Db 1022 TLNDGGLSIKNTPTGSEQIQVGADGVKFAKVNNGVVGAGIDGTTRITRDEIGFTING-S 1080
QY 816 LHQDAAMDVRMPEGSFQSPATKEEYTNV-----VIANNVDFKFSWGITDEM-APQY 870
Db 1081 LDKSKPHLSK--DGIN-----AGGKITNIOGSEIAQNSHDAVGTGGKIYDLKTELENK 1131
QY 871 VST-----DQQLDSVIQNGYAFTRDYDLGMSK-----ANKYTAQOLVK 911
Db 1132 ISSHTAKTANSLHFSVADEQGNFTVSNPYS---SYDTSKTSDDVTTFAGENGITTKVKN 1188
QY 912 AIKAL---HAKGLKVMADWVPQMYTFPKQEVVTVTRTDKFGKPIAGSQINHSLYYTD-- 966
Db 1189 GVVVRGIDDTKGL-----TTPK-----LTVGNNGKGIVIDSQNGQNTITGLS 1231
QY 967 -----TKSSGDDYQAKYGAFDELKEKYPELFTKKQI-----STGOAIDPSVKI KQWS 1015
Db 1232 NTLANVTNDKGSVRTTEQNIKIDEDKTRAASIVDVLSAGFNLOQNGEAVD---FVSTYD 1288
QY 1016 AKYFNGSNILGRGADVYLDQVSNKYP--NVASDTLFLPSSLLGKVVESGIRVDGK---1070
Db 1289 TVNFDAGNATTAKTVYDDTSKTSKVYDVNVVDTTIEVKDKKLG-VKTTTLTSTGTGANK 1347
QY 1071 YIYNSSATGQ-VKASFI-----TEAGNLYYFGKQGYWMTGAQTINGANYFFLENGTALR 1124

Qy	1149	-YQFQGNDRYFKD-----GNMAVGLTTVD-----GNVQYFDKD-GV-----QAKD	1187
Db	1055	NYVRTNDGLAFNDASAQGVGATAIGNSVAKGSSVAIGQGSYSDVTGIALGSSSVSS	1114
Qy	1188	KII-----VTRDGKRYRFDONGNNAATNFTADKTGHWYILG-KDGVAVTGAQTVGK	1238
Db	1115	RVIAKSRDTSITENGWVIGYDITDGBELLGALSIGDDGKQRIINIVADGSEAHDAVTVRQ	1174
Qy	1239	QK-----LYFRANGQOVKGFVTSDEGKLYFYDVDSGD-----MMWDTFIEDKAG	1283
Db	1175	LQNAIGAIVATPTKYFPHANSTEEDSLAVGTDLSLAMAQKTIWNGDKGIGYGVAVDANAL	1234
Qy	1284	NWFLYLGKDGRAV-----TGAQTIIRGOKLYF-----KANGQOVKG	1317
Db	1235	NGIAGSNAQVIHNSIAIGNSTTTTGAQNTYTAYNMDAPONSVGSEFSGVSGADGQRIIT	1394
Qy	1318	DIVKG---TD---GKIRYDIAKSGEQV---FNKTVKAAADGKTYVIGNDGVAVDPSPV	1365
Db	1295	NVAAGSADTDVAVNGQLKVTDQAQVSQNTQSTITLNDNRVTNLSRVNTIEN---GIGDIVT	1351
Qy	1366	KGQT--FK-----DASGALRFPYNLKQLVTGSGWYETANHDIWYIOSGKALTGEOTI--	1415
Db	1352	TGSTKYFKTNTDGVDAASA---QKQDSVAIGSGSIAAADNS-VALGTGSVATEENTISV	1405
Qy	1416	---NGOHLFYKEDGHQVK---GQLVTGTDGKRVRYDANSBGDQAFNKSVTVNGKTYIF	1466
Db	1406	GSSTNQRRIITNVAAGKNATDAVNVAQLKSSAGEVR-YDTKAGSDIDYSNITLGG----	1459
Qy	1467	GNQTAQTAGNPKG-----QTFKDGSQDIRFYSMEGQL-----	1498
Db	1460	GNQGTTRISNVAGVNNNDVVNYAQLKQSQVETKQYTDQRWEMDNKLSKTESKLGGIA	1519
Qy	1499	-----VTGSGWYE-----NAQOGWLTVKNG	1518
Db	1520	SAMAMTGLPQAYTFPGASMAISIGGTTNGESAVALGVSMVSGANGRWYFKLG	1570

RESULT 11

```

US-10-793-626-2964
; Sequence 2964, Application US/10793626
; Publication No. US2005025478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2964
; LENGTH: 5024
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (5024)
; OTHER INFORMATION: variable amino acid
; US-10-793-626-2964

```

Qy	80	TDAATDOASABOTQGTAS-----TDTAAQTNNANEAKWP-----	117
Db	2466	AEAKSDGNOGIEHLOGLTOSQDKQHQDLINQAOTKQOVDDIVNNSKOLDSMNQLQOIVN	2525
Qy	118	---TENENQGTDB-----MLAEAKGVATAESDSTPSDLAK-----MSNVKQVD	158
Db	2526	NDNTVKQNSPFINEDSSQQDAYNHAQAAKDLITAH-----PTIMDKNOIDQAIENIKOAL	2581
Qy	159	GKYYYYQDQGNVKFNFAVSGDIIYFDE--TGAYKDT-----SKVDADKSSAV	206
Db	2582	NDLHGSNKLSEDKX---EASEQLQNLISLTQKOTILNHIFSAPTRSQV--GEKIASAK	2636
Qy	207	SQNAITFA-----ANRAVYSTAKNF-----EAVDNYLTADSVWRPKSIILKDGKTWTESK	257
Db	2637	QLANNTKALRDSIADNNEILQSSKYFNEDSEQNAYNOAVN--KAKNIINDOPT--PVMAN	2693
Qy	258	DDFRPILMAWMPDETETKRNYYN--YMKNVGIDKITYTAETSOADLTAAAEVLQAR---IBQ	313
Db	2694	DETQSVL-----NEVQTKDNLHGOOKLANDKT-----DAQATLNALNYLNOAQRGNLET	2743
Qy	314	KITSENNKTWLRBAISAFVKRTPQWNGESEKPYDDHLQ--NGALL-----F	357
Db	2744	KVQNSNS---REVQKVQLANQLN--DAMKKLDLALTCNDALIKOTSNYINEDTSQOVNF	2798
Qy	358	DNQDTLPTQOSNYRLNFTPTNQSGLSRFTYFNPNDPJGGVDFPLANDVNSNPV---414	
Db	2799	DEYTRGKNIVABQTPNPNMSPTN--INTIADKITEAKND--LHGQVKLEQAQOQSINTINQM	2856
Qy	415	-----VOAEQOLNW-----LHYLLNFGSIYANDADANPDSIRVDADVNDLLOI	459
Db	2857	TGLNQAKQELNOEIOOTTRSEVHOVI-----NKAQALNDSMNTLRQSTDBEHVKQ	2909
Qy	460	SSDYLKAAAGIDIKKNNKANNHVSIVBAWSDNPTPYLHDDGDLNMNDKFRLSMLWSLAK	519
Db	2910	TSNYINETVGNQAYNNAVDVRVKQIINQTSNPT-----MN-----	2944
Qy	520	PLD--KRSGLN-----PLIINSLVDRVDDREVETVPSYFARAHSEVODIIRDIKABI	573
Db	2945	PLEVERATSNVKTSDALHG--ERELDNK-----NSKTFVANHLDNLNQAQKEALTHEI	2997
Qy	574	NPNSPGYSFTQEBIEQAFIYNE-----DLKK--TDKKYTHYVNPVPSYLLTLTNKGSIPR	626
Db	2998	EQATI-----VSQVNNIYNKAKALNNDMKKLDIVAQQONVRQSNNYI--NEDSTPQ	3047
Qy	627	VYXGDMFTDDG---QYMAKNTVNYDAIESL---LKARKMYVSGGQAMQNYQIENGEBILTS	680
Db	3048	NMYNDIINHAQSIIDQVAPNTWSHDEIENAINNIKHAINALDGEHKLQOAKENAMLLINS	3107
Qy	681	VRYKGKALKQSDK---GDATTRTSGVVMGNQPNFSLDGKVVALNMGAAHANQBYRALM	737
Db	3108	LNDLNAPORDAINRLVNEAQTREKVAEQLOSAQ-----ALNDAMKHLRNSIQNS	3157
Qy	738	VSTKDGVAIYATADAKAGLVKRTDENGILYFLND-----DLKGVANPOV	783
Db	3158	SVQESKYINASDAKEQYNHAVREVEN---IINEQHPTLDKEIIKQLTDAVNOANNDL	3213
Qy	784	SGFLQVWVPVGAADDQDIR-----VAASDTASTDGSLSHQD	819
Db	3214	NG-----VELLDADKQNAHQSIPTLMHLNQAOONALNEKINNNAVTRAKVAIIQOAKIIL	3268
Qy	820	AAMDS-----RVMPEGSNF--QSFA TKBEBYTNV-----IANNVDKVFVSWGITDPE	865
Db	3269	HAMENLEESIKOEQVQKSSNYINEDPPOQETYNNAVDHVTEILNQTVNPTLS--IEDIE	3326
Qy	866	MAPQYSSDTGGQFLDSVIONGYAFTRDYDLGHSKANKYGTADOLVKATKALHAKGLKWA	925
Db	3327	HAINEVNQAKKQI-----RGKQKLYQTID-----LADKELSCLD	3360
Qy	926	DWPDQMYTFPPQEVVYVTRTDKFGKIAGSOIHNHSLVYTDTKSSGDDYQAKYGGAFIDE	985
Db	3361	DLTSQOSSISINQIYAKTRTEVAQAEKAKSINHAM-----KALNKIY--KNADKVLDS	3413
Qy	986	---LKEYPELEFTKKQI-----STGQAI DPSPV-----K	1010


```

; SOFTWARE: FaastSEQ for Windows Version 4.0
; SEQ ID NO 228
; LENGTH: 396
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-11-022-562-228

Query Match      3.1%; Score 257; DB 7; Length 396;
Best Local Similarity 25.2%; Pred. No. 3.7e-07;
Matches 131; Conservative 64; Mismatches 151; Indels 174; Gaps 30;

QY 969 SSGDDYQAKYGGAFDELKEKYPFLTKKOISTGQAIDPSVKIK--QWSAKYFNGSNI-- 1024
DB 25 SSSFYKWSSTEGS--DFILVRYLESNKK-----ILQKIRIKGLSTKSFNKMIDF 75
QY 1025 -----LGRGADYVLSDOVSNKYFNVASDTLFLPSSLLG-KVYESGIRYDGKGYIYNSSAT 1078
DB 76 KDIKKLSLG--YIMS--NFKSFENSE--LDRDHLGPKII-----DNKTYVYDEAS- 120
QY 1079 GDOVKASFITEAGNLYYFGK-DGYMVTGAQTINGANYFPLENGTALRNTIYTDAGNSHY 1137
DB 121 --KLVKGLININNSLFYDPDIESNLVTGMQTINGKKYF-DINTGAASTSYKIING-KHF 176
QY 1138 YANDGKRYEYGYQQFG-----NDWRYFKDGNMAVGLTTVDGNVQYFDKDGVOAKDKIIVT 1192
DB 177 YFNN-----NGVMQLGVFKGPDGFYFAPAN-----TONNNI-----EGQAIYVQSKFL-T 221
QY 1193 RDGKRYRFDQHNHNAATNTFTIADKTGHWYLYLGKGVAVTGAQTVGKQKLYFANGQQVKG 1252
DB 222 LNKGYFYD-----NDSKAVTGMQITDGKKYFNLNTAEAAAT 258
QY 1253 DFTVSDGKLYPYVDSDGDMWTDFTIEDKAGNWFYLGKDGAAVTGAQTIRGOKLYFKANG 1312
DB 259 GQWTTID-GRKYYFN-----TNTSI-----ASTGYTIINGKHFFYFNTDG 295
QY 1313 QQVKGDIKVGTDGKIRYDAKSGEQVFNKTVKAADGKTVYIGNDGVAVDPSPVVGQTFKD 1372
DB 296 IMQIG-VFKGPG-PEYF-----APANT-----DANNIEGQ----- 324
QY 1373 ASGALRFYNLKQLVTGSGWYETANHDWYI--QSGKALTGEQTINGOHLFYFKEDGHQVK 1430
DB 325 ---AIRYQNRFLYL-----HDNIYFNGNSKAVTGMQTINGNVVYFMPD----- 365
QY 1431 GQLVTGTDGKRYDYDANSQDQAFNKSVTVNGKTYTFGNDG 1470
DB 366 -----TAMAAAGGLFEIDGVIYFPGVDG 388

RESULT 13
US-10-873-528-17
; Sequence 17, Application US/10873528
; Publication No. US20050276814A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129WO
; CURRENT APPLICATION NUMBER: US/10/873,528
; CURRENT FILING DATE: 2004-06-23
; PRIOR APPLICATION NUMBER: US/09/769,787
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 658
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-873-528-17

3414 SRFENDOPEKAYQQAIIINHVSIIHROTPENMDPTVINSIITHELETAQNHLHGQKLAH 3473
QY 1011 IKQWSAKYFNGSNILGRGADYVLSQOVSNKYFNVASDTLFLPSSLLGKVVESGIRYDGKG 1070
DB 3474 AKQDAANVINGL-----IHLNVAQREVMIWT----- 3499
QY 1071 YIYNSATGDQVKASFITAGNLYYFGKGYMVTGAQTINGANYFPLENGTALRNTIYTD 1130
DB 3500 ---NTNATTREKVAKNLDAQAL-----DKAMET-----LQGVAAHKNILND 3539
QY 1131 AQGNSHYANDGKRYEYGYQQFGNDWRYFKDGNMAVGLTTVDGNVQY---PDKDGVQAKD 1187
DB 3540 SK-----YLNEDSKYQQQYD-----RVTADEAQLNQTPTLPEYKVDIVKDNVLANE 3588
QY 1188 KIIIVTRDGRVYFDQHNHNAATNTFTIADKTGHWYLYG-----KDGVAVTGAQTVGKQK 1240
DB 3589 KILFGAE-KLSY-DKSNAN-----DEIKHMYLNNAQKQSIKDMISHAALRTEVKQL 3638
QY 1241 LYFEANGQQVKGDFVTSDEGKLYFYDSDGDMWTDFTIEDKAGNWFYLGKDGAAVTGAQT 1300
DB 3639 L-----QQAK-----TLDEAMKSLD-KTQVVTIDTTLF-----NYTEASEDKKEKVDQTV 3683
QY 1301 IRGOKLYFKANGQQVKGDIK-----GTGDKIRYDAK----- 1333
DB 3684 SHAQALIDKINGSNLSLDQVRQALEBQTOASENLDDGQVRVEAKVHANOTIDOLTHLSL 3743
QY 1334 ---SSEQVFNKT-----VKAADGKTVYIGNDGV 1358
DB 3744 QQTAKESVKNATKLEIATASNALALNKMVKLEQFINHADS IENSNDYRQADDKII 3803
QY 1359 AVDPSPVVKQ-----TFKDGASGALRFYNLKQOLV---TGSGWYETANH-----DWYIQT 1404
DB 3804 AYDDALEHGQDIQKSNATQNEAKQALQ-----QLINAEATSLNGFERLNHARPRALEYIK 3857
QY 1405 S-----GKALTGEQTINGOHLFYFKEDG---HQVKGQLVG-----TDGKV 1441
DB 3858 SLEKINNAQSALEDKVTQSHDLLEHLV--NEGTLNLDIMGELANAIWNVYAPTAKSI 3915
QY 1442 RYDA-----NSGDOAFNKSVTVNGKTYVFGNDGTAQ----- 1473
DB 3916 NYINADNLKDNFTQAINARDALNT---OQQLDFNAIDTFKDDIFKTKDALNGIERL 3972
QY 1474 TAGNPKGQTFKQSGDIRFY-----SMEQQ 1497
DB 3973 TAAKSAEKLIID--SLKFINKAQFTHANDEIMNTNSIAQLSRIVNOAFDLNDAMKSLRDE 4030
QY 1498 L-----VTGSGWYENAO-----GQWLYVKNKGKVLTLGQTVGSGORVYFDE- 1536
DB 4031 LNNQAHVPQASSNYINSDLEDLKOQFDHALSNARKVLAKENGKRLDEIQIEGLKQVIEDTK 4090
QY 1537 ---NGIQ-----AKGKAVR 1547
DB 4091 DALNGIQRLSKAKAKAIQ 4108

RESULT 12
US-11-022-562-228
; Sequence 228, Application US/11022562
; Publication No. US20050249742A1
; GENERAL INFORMATION:
; APPLICANT: Ruprecht, Ruth M.
; APPLICANT: Shisong, Jiang
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING
; TITLE OF INVENTION: A CYTOTOXIC T LYMPHOCYTE IMMUNE RESPONSE
; FILE REFERENCE: DFN-043CN
; CURRENT APPLICATION NUMBER: US/11/022,562
; CURRENT FILING DATE: 2004-12-22
; PRIOR APPLICATION NUMBER: PCT/US03/20322
; PRIOR FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: 60/392718
; PRIOR FILING DATE: 2002-06-27
; NUMBER OF SEQ ID NOS: 340
```

```

Db      47  WQGRQYLKEDGS-----QAAENWFPDTHYQSWFYIKADANAENWLKQ-----GDD  93
Qy      1071 YIYNSSATQOVKASFITEAGNLYYFKGQGYMTGAQTINGANYFFLENGTALRNTIYTD  1130
Db      94  YFYLKSG-GYNKSEWVEDGAFYFLDQDGKRNANW--GTSYVGATGAKVITEDWY-D  149
Qy      1131 AQGNSHYYAN-DGKRREYNGYQQF-GNDWRYFKDGNMAVGLTTVDGNVQYFDPKDGVOAKDK  1188
Db      150 SQYDAWFYIKADGQHAKEWLOIKGDY-YFKSGYL--LTSQWINGAYYNASGAKVQQG  206
Qy      1189 IIVTRDGKRYFDOHNGNAATNTFIADTKGHTYFLKGDGVAVTGAQTGKQK-LYFEANG  1247
Db      207 WLFDPKQYQSWFYIKENGYADKEWIFE-NGHYTYLKSGGYMAANETWIDKESWFLYKFDG  265
Qy      1248 QVQKGDFTVSDEGLKXPYDVDSGDMWTDTFIEDKAGNWFYLGKDGAAVTAQOTIRGQ--  1305
Db      266 KMAEKWYDSSHQAWTYFFKSGGYMTANETWIDKE-SWFLYKSDGKIAKEWYVDSSHQA  324
Qy      1306 -LYFKANGQQVKGDIV-----KGTGDKIR-----YYDAKSGQGVFNKTV  1343
Db      325 WYTFKSGGYMTANETWIDKESWFLYKSDGKIAKEWYVDSSHQAWTYFFKSGGYMAKNETV  384
Qy      1344 KAADGKTYVIGNDGVAVDPSPVVKQGTFKDASGALRFPNLKQLGVTSQGWYETANHDMVYI  1403
Db      385 ---DG---YOLGSDG---KWLGGKTTNENAA---YYQV---VPVTANVYDSDGEXLSYI  428
Qy      1404 QSGKALTGEQ-----TINGOHLFYKPDGHQVKGQVLVTGTDGKRVYYVDANSQDQ-  1451
Db      429 SQGSVWMLDKRKSDDKRLAITISLGSYMYKTEDLQA---LDASKDFIPYIESD-GHRP  483
Qy      1452 ----AFNKSVTVN-----GKTYTFGN-----DGTAGTAPNGKGTFFKDGSDIRFYSME  1495
Db      484 YHYVAQNASIPVASHLSMEVGKYYISADGLHFDGF--KLENP--FLFKDLTEATNYSAE  539

RESULT 15
US-10-793-626-1780
; Sequence 1780, Application US/10793626
; Publication No. US20050255478A1

```

```

; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1780
; LENGTH: 1155
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
; US-10-793-626-1780

Query Match 3.0%; Score 249.5; DB 6; Length 1155;
Best Local Similarity 19.4%; Pred. No. 4.2e-06;
Matches 293; Conservative 173; Mismatches 494; Indels 553; Gaps 75;

Qy 1 MEKNVRFRKMHVKKK-----RWVTLSSVAS-----ATMLASALGASVASAD----- 39
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 9 INKRVDFLSKNVKYSIRKFTVGTASILVGTALMFGCAADNEAKAEDNQLSASKEEQKG 68
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

```

QY	40	-TDTASDDSNQAV-----TGQDTN-NOATDQTSIAATATSEQASGASTDAATDQASAAEQ	92
Db	69	SRDNESKLNQVDLDNGSHSEKTTNNVNNATEVKKVEAPTTS-----DVSKPKANEAVV	122
QY	93	TGQTTASTDTAAQTNTTNAEAKVVPTEINENQGFTEMLAEAKNVATASDSITPSDLAKMS	152

Db 123 TNESTKPKTTEA-----PTVNE-----ESIAETPKTSTTQODSTEKNPSLK 164
Qy 153 NYKQVDGKYYYYDQGNVKNFAVSGDKIYYFDETGAYKDTSKVDADKSSSAVSQNAI 212
Db 165 D-----NLN5STTSKESKTDHSTKQAQMSTNKSNDLNDSPQSEKTS 209
Qy 213 FRANNRAYSTSAKNFAVDNYITADSWYRPK-----SILKDGKTWTESGKDD 259
Db 210 SQANNDSTNQAPSQKLDSPSEQVKYTKFNDPEPTQDVEHTTKTKLTPSISTDSSVND 269
Qy 260 FRPLLMAWPDTEKBNVYNNKVVGDIKTWTAEYSQA-----DLTAAAEVLQARIQOK 314
Db 270 -----KODYTRSAVASLGVD5NETEALITNAVRONLDLKAAS-----REGINEA 312
Qy 315 ITSENNTKWLREAI5AFVKTQPMNGESEKPYDDHLLQNGALLFDNQDTLTPDTQSNYR-- 372
Db 313 IIAEALKK-----DFSNP-DYGVDTPLALNTSQSKNSPHKSPRWN 353
Qy 373 LLNRTPTNQTGLSDSFTYNPDPLGGYDFLLANDVDNSNPVVQABQLNWLHVLNPGSI 432
Db 354 LMSLAAEPNSGK-----NVND-----KVKITNPTLS-----LNKSN 385
Qy 433 YAND-----ADANFDSIRVDADVNDVADLLQISSDYLKAAYGIDKKNKNANNHVSIVEAW 487
Db 386 HANNVIWPTSNOFN-----LKANYELDDSIK----- 412
Qy 488 SOND-----PYLHDDGDNLMNMNKPRLSMLSLAKPLDKRSGLNPLIHNSLVDREYD 541
Db 413 -EGDTFTIKYGYIRPGGLEPAIKTQLR-----SKDG-----SIVANGVY 452
Qy 542 DREVEYVPSYSPARAHSDSEVDIIRDIKABINPNSFGYSFTQEBIEQAPKYNEDLKKI 601
Db 453 DKTNTT-TYTFNTYVD-QYQNI-----TGSF-----DLIAT 482
Qy 602 DKYKT-----HYNVPLSYTLTLLTNKSGI PRVYVGMFTDDQYMANKTVNYDA-IESLLKA 656
Db 483 PKRETAIKDNQNYPMEVTI-----ANEVVKDFIVDYGKNKDNITTAAVANVDNVNKK 535
Qy 657 RMKYVSGGQAMQYQIGNGEILTSVRYGK--GALKQSDKGDAIT-----RTSGVG 704
Db 536 HNEVVYLNQNNQPKY--AKYFSTVNGKFIPEGVKVYEVTDINAMVDSFNPDLNSSNVK 593
Qy 705 VVNGN--QPNFSLDGKVVALNM-----GAAHAN 730
Db 594 DVT5QFTPKVSADGTRVDINFARSMANGKKYIVTQAVRPTGTGNVYTEYWLTRDGTNTN 653
Qy 731 QEYRALMVSTKDGVA7YATDADASKAGLVKRTDENGYLFLNDDLKGVANPQVSGFLQYW 790
Db 654 DFR-----GTKSTVTYLN5SSTAQGD--NPTYSLGDYVWLDKKNKGVDQDDDEKGLAGVY 707
Qy 791 VPVGAADDQDIRVAASDTASTDGKSLHQDAAMD5RYMFEFGSNFQS-----PATKEEY 844
Db 708 VTLKDSNNRELQ-----RVTTDQSGHYQ-----FDNLQNGTVTVEFAL-PDNY 749
Qy 845 TNVVIANNVDKVF5W-----GITDFEAPQYV5STGQFLDSVIQNGYAFTRDYDLG--- 896
Db 750 T5PANNSTND5IDSDGERDGRKVVVAKGTINNADNMTVDT-----GFYLTPEKYNVDYV 805
Qy 897 MSKANKYGTADQLVAKAIKALHAKGLKVMADWPDQYTFPKQ5VTVTRTDKFGKPIAGS 956
Db 806 WEDTNKGIGQDDNEKISN-----VKVTLKNKNGDTIG-- 838
Qy 957 QINHSLYVTDTSKSGD-DYQAKYGGAPLDELKEKYPELFTKKQISTGOAIDPSVKIKQWS 1015
Db 839 -----TTTDD5NGKVEFTGLENGDYTIEFET--PEGYTPTKQNSGD-----E 879
Qy 1016 AKYFNG--SNILGRGADYVLSQV5KNYFNVASDTLFLPSSLLGKVVESGIRYDGKY-- 1071
Db 880 GKDSNGTKT7TVTKDAD-----NKTIDSGF-----YKP 907
Qy 1072 IYNSSATGQOVKASFI5EAGNL5YFGKDGVMVTGAOTINGANYFFLE-NGTALRNTIYTD 1130
Db 908 IYN---LGD-----YVWEDTN-----KDGIQDDSEKISGKVKVTLKDKNGNAI-GTTTID 953

Qy 1131 AOGNSHYVANDGKRYENG5YQQFNGNDWRYFKDGNMAVGLTTVDGNVQYFDKDGVOAKDKII 1190
Db 954 ASG--HY---QPKGLENGSY-----TVE----- 971
Qy 1191 VTRDGKRYFDOHNGNAATNTFIADKTGHWYVLGKDGVAVTGAQTVGKQKLYFEANGQQV 1250
Db 972 -----FETPSGYTPTKA-----NSGQDITVDSNGITTTGI-----ING--- 1004
Qy 1251 KGDFVTSDEGKLYFYDVIDS---GD-MWTDFTIEDKAGNWFYLGKDGAAVTVGAQTIRGQKL 1306
Db 1005 -ADNLTIDSG---FYKTPKYSVGDYVWEDT-----NKDGIQDDNEKISGKVKV 1048
Qy 1307 YPKANGQQVKGDI5VKG---DGKIRYVDKSGE5QVFNKTVKAA5DKT5VIGNDGVAVDP 1362
Db 1049 TLK-----DEKGNIIISTTTTIDENGKYQFONLDSNGYIIH--PEKPEGMTQTTANS 1097
Qy 1363 SVVKGQTPK5DASG 1375
Db 1098 -----NDDEK5DADG 1106

Search completed: February 11, 2006, 20:58:23
Job time : 22.048 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 19:15:50 ; Search time 32.8183 Seconds
(without alignments)
4661.567 Million cell updates/sec

Title: US-10-797-821-37

Perfect score: 8349

Sequence: 1 MEKNVRFKHKVKKRWLTLS.....VGYQYFNGDGARIYRGWN 1590

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 80.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8256	98.9	1592	2 A38175	glucosyltransferas
2	4580	54.9	1475	2 B3135	gtfB protein precu
3	4214.5	50.5	1375	2 J70345	dextranucrase [EC
4	3599	43.1	1431	2 A45866	dextranucrase [EC
5	3325.5	39.8	1518	2 A4811	glucosyltransferas
6	3245	38.9	1577	2 T30858	glucosyltransferas
7	3214.5	38.5	1599	2 S22737	glucosyltransferas
8	3128.5	37.5	1449	2 T30857	glucosyltransferas
9	3114.5	37.3	1449	2 T30552	glucosyltransferas
10	3088	37.0	1508	2 T31098	probable dextranu
11	3031	36.3	1365	2 A41483	glucosyltransferas
12	3021	36.2	1290	2 JCS473	dextranucrase [EC
13	678	8.1	2817	2 B97033	uncharacterized pr
14	518	6.2	563	2 A37184	glucan-binding pr
15	513	6.1	2710	2 A37052	toxin A - Clostrid
16	435.5	5.2	648	2 S10869	enterotoxin A - Cl
17	416	5.0	2178	2 S55805	alpha-toxin - Clo
18	400.5	4.8	2364	2 I40884	cytotoxin L - Clo
19	388	4.6	2366	2 S10317	toxin B - Clostrid
20	381	4.6	2367	2 S70172	toxin B - Clostrid
21	329	3.9	1806	2 AF1717	probable peptidogl
22	328.5	3.9	2334	2 S32920	cell wall-associat
23	311	3.7	1385	2 D89824	hypothetical prote
24	299.5	3.6	1829	2 E81086	iron-regulated pro
25	298.5	3.6	1946	2 AE1449	hypothetical prote
26	291.5	3.5	1829	2 S35027	cytotoxin Rtx homo
27	291	3.5	1315	2 T28679	fibrinogen-binding
28	284.5	3.4	1463	2 T30290	AAS surface protei
29	277	3.3	2167	2 AF1489	cell wall-associat

ALIGNMENTS

RESULT 1

A38175

Glucosyltransferase precursor - Streptococcus sobrinus

C:Species: Streptococcus sobrinus

C:Date: 28-Aug-1992 #sequence_revision 28-Aug-1992 #text_change 15-Mar-2004

C:Accession: A38175

R:Abo, H.; Mtsamura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.

J. Bacteriol. 173, 989-996, 1991

A:Title: Peptide sequences for sucrose splitting and glucan binding within Streptococcus

A:Reference number: A38175; MUID:91123227; PMID:1704006

A:Accession: A38175

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1592 <ASO>

A:Cross-references: UNIPARC:UPI000012BCB2; GB:D90213; NID:G217032; PIDN:BAAL4241.1; PID

F:1093-1112/Domain: cpl repeat homology <CP1>

F:1222-1241/Domain: cpl repeat homology <CP2>

F:1287-1306/Domain: cpl repeat homology <CP3>

F:1330-1351/Domain: cpl repeat homology <CP4>

F:1352-1371/Domain: cpl repeat homology <CP5>

F:1402-1420/Domain: cpl repeat homology <CP6>

F:1465-1484/Domain: cpl repeat homology <CP7>

F:1513-1532/Domain: cpl repeat homology <CP8>

Query Match

Best Local Similarity 98.9%; Score 8256; DB 2; Length 1592;

Matches 1574; Conservative 6; Mismatches 10; Indels 2; Gaps 1;

QY	1	MEKNVRFKHKVKKRWLTLSVASATMLASALCASVASADTDTSDDSNQAVVTGDDTTNN	60
DB	1	MEKNVRFKHKVKKRWLTLSVASATMLASALCASVASADTDTSDDSNQAVVTGDDTTNN	60
QY	61	QATDQTSIAATATSQSASTDAATDQASAAEQGTGTTASTDTAAQTNNANEAKVPTEN	120
DB	61	QATDQTSIAATATSQSASTDAATDQASAAEQGTGTTASTDTAAQTNNANEAKVPTEN	120
QY	121	ENQGTDEMLAEAKNVATAESDPSDLAKMSNVQVDGKYKYDDQGNVKNFAVSVDG	180
DB	121	ENQGTDEMLAEAKNVATAESDPSDLAKMSNVQVDGKYKYDDQGNVKNFAVSVDG	180
QY	181	KIYFDETGAYKDTSKVDADKSSSAVSONATIFAANNRAYSTSAKNFEADVNYLTADSWY	240
DB	181	KIYFDETGAYKDTSKVDADKSSSAVSONATIFAANNRAYSTSAKNFEADVNYLTADSWY	240
QY	241	RPKSLKDGKWTESGKDDFRPLMAWPDFTETKENYNNMKVGVGIDKTYTAETSQADL	300
DB	241	RPKSLKDGKWTESGKDDFRPLMAWPDFTETKENYNNMKVGVGIDKTYTAETSQADL	300
QY	301	TAAAEVQARIEQKITSNNTKWLRSAISAFVKTPQWNGESEKPYDDHLQNGALLFDNQ	360
DB	301	TAAAEVQARIEQKITSNNTKWLRSAISAFVKTPQWNGESEKPYDDHLQNGALLFDNQ	360

probable peptidogl
lactocepin (EC 3.4
endo-beta-N-acetyl
surface protein XF
probable adhesin Z
probable adhesin E
hypothetical prote
hypothetical prote
endo-beta-N-acetyl
hypothetical prote
dextranase inhibito
iron-regulated pro
hypothetical prote
hypothetical prote
surface protein ps
surface protein ps

QY 361 TDLTPDTOSNYELLNRTPTNQTGSLDSBPTYNPNPDLGGYDPLLANDVDNSNPVQABQL 420
DB 361 TDLTPDTOSNYELLNRTPTNQTGSLDSBPTYNPNPDLGGYDPLLANDVDNSNPVQABQL 420
QY 421 NWLHYLLNFGSIYANDADANFDSIRVDAVDNVDADLLOI SSDYLKAAAYGIDKNNKANNNH 480
DB 421 NWLHYLLNFGSIYANDADANFDSIRVDAEDNDVADLOI SSDYLKAAAYGIDKNNKANNNH 480
QY 481 VSIWEASNDNTPYLLHDDGDNLMNNDKFRLSMLWSLAKPLDKRSLGNPLIHNSLVDRV 540
DB 481 VSIWEASNDNTPYLLHDDGDNLMNNDKFRLSMLWSLAKPLDKRSLGNPLIHNSLVDRV 540
QY 541 DDRETVTPSYSFARAHSEVODIIRDIKABINPNSFGYSTOREIEQAQFIYNEDLKK 600
DB 541 DDRETVTPSYSFARAHSEVODIIRDIKABINPNSFGYSTOREIEQAQFIYNEDLKK 600
QY 601 TDKKYTHNVPLSYLLLTNKGSIIPRVYVYGMFTDDGOYMAKNTVNYDAIESLLKARMKY 660
DB 601 SDKKYTHNVPLSYLLLTNKGSIIPRVYVYGMFTDDGOYMAKNTVNYDAIESLLKARMKY 660
QY 661 VSGQAMQNYQIENGSEILTSVRYGKALQKSDGDAITRTSGVGVMMGNQPNFSLDGKV 720
DB 661 VAGQAMQNYQIENGSEILTSVRYGKALQKSDGDAITRTSGVGVMMGNQPNFSLDGKV 720
QY 721 ALNMGAAHANOBYRALMYSTKDGVAATYATDADASKAGLVKRTDENGILYFLNDDLKGVAN 780
DB 721 ALNMGAAHANOBYRALMYSTKDGVAATYATDADASKAGLVKRTDENGILYFLNDDLKGVAN 780
QY 781 PQVSGFLOVPVGAADQDDIRVAASDTASTDGKSLHODAAWDSRMVFEFSGNFOSFATK 840
DB 781 PQVSGFLOVPVGAADQDDIRVAASDTASTDGKSLHODAAWDSRMVFEFSGNFOSFATK 840
QY 841 BEEYTNVVIANNVDFXVSWGITDFEMAPQYVSSSTDGQFLDSVIQNGYAFTRIDYDLGMSKA 900
DB 841 BEEYTNVVIANNVDFXVSWGITDFEMAPQYVSSSTDGQFLDSVIQNGYAFTRIDYDLGMSKA 900
QY 901 NKYGTADOLVKAIKALHAKGLKWMADWPDQMYTFPKQEVVTVTRTDKFGKPIAGSQINH 960
DB 901 NKYGTADOLVKAIKALHAKGLKWMADWPDQMYTFPKQEVVTVTRTDKFGKPIAGSQINH 960
QY 961 SLYVTDTKSSGDDYQAKYGGAFDLDELKEKYPBLFTKKOISTGOAIDPSVKIKQWAKYFN 1020
DB 961 SLYVTDTKSSGDDYQAKYGGAFDLDELKEKYPBLFTKKQWSTGOAIDPSVKIKQWAKYFN 1020
QY 1021 GSNILRGADYVLSDOVSNKYFNVASDTLFLPSSLGLKGVESGIRYDGKGIYNSSATGD 1080
DB 1021 GSNILRGADYVLSDOVSNKYFNVASDTLFLPSSLGLKGVESGIRYDGKGIYNSSATGD 1080
QY 1081 QVKASFITEAGNLVYFGKDGVMVTAQOTINGANYEFLENGTALRNTIYTDAGNSHYAN 1140
DB 1081 QVKASFITEAGNLVYFGKDGVMVTAQOTINGANYEFLENGTALRNTIYTDAGNSHYAN 1140
QY 1141 DGKRY - - ENGYQOFGNDWRYFKDGNMAGVLTVDGNVOYFDKDGVOAKDKIIVTRDGKVR 1198
DB 1141 DGKRYENENGYQOFGNDWRYFKDGNMAGVLTVDGNVOYFDKDGVOAKDKIIVTRDGKVR 1200
QY 1199 YFDOHNGNAAVNTFTIADKTGHWHYLGKDGVAVTGAQTVGKQKLYPEANGQQVKGDFVTS 1258
DB 1201 YFDOHNGNAAVNTFTIADKTGHWHYLGKDGVAVTGAQTVGKQKLYPEANGQQVKGDFVTS 1260
QY 1259 EGKLYFYDVSDDMTDTFIEDKAGNWFYLGKDGAAVTAQOTIROKLYFKANGQQVKG 1318
DB 1261 EGKLYFYDVSDDMTDTFIEDKAGNWFYLGKDGAAVTAQOTIROKLYFKANGQQVKG 1320
QY 1319 IVKGTGDKIRYVDAKSGQVFNKTKAADGKTYVIGNDGVAVDPSPVWKQGFKDGASGALR 1378
DB 1321 IVKGTGDKIRYVDAKSGQVFNKTKAADGKTYVIGNDGVAVDPSPVWKQGFKDGASGALR 1380
QY 1379 FYNLKGQVLTGSGWYETANHDWVYIQSGKALTGEQTINGOHLYPFKEDGHQVKGQVLTGTD 1438
DB 1381 FYNLKGQVLTGSGWYETANHDWVYIQSGKALTGEQTINGOHLYPFKEDGHQVKGQVLTGTD 1440
QY 1439 GKRYRDANSQDAQFNKSVTVNGKTYIFGNDGTAQTAGNPKGQTFKDGSDIRFYSMEGQL 1498

DB 1441 GKRYRDANSQDAQFNKSVTVNGKTYIFGNDGTAQTAGNPKGQTFKDGSDIRFYSMEGQL 1500
QY 1499 VTGSGWYENAOQWLYVKNKGKVLTLGLQTVGSQRYVYFDENGIOAKGKAVRTSDGKIRYFDE 1558
DB 1501 VTGSGWYENAOQWLYVKNKGKVLTLGLQTVGSQRYVYFDENGIOAKGKAVRTSDGKIRYFDE 1560
QY 1559 NSGSMITNWKVYFYGOYYFYFGNDGARIYRGWN 1590
DB 1561 NSGSMITNWKVYFYGOYYFYFGNDGARIYRGWN 1592
RESULT 2
B33135
gtfB protein precursor - Streptococcus mutans
C:Species: Streptococcus mutans
C>Date: 23-Oct-1990 #sequence_revision 23-Oct-1990 #text_change 09-Jul-2004
C:Accession: B33135; A33128
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
J. Bacteriol. 169, 4263-4270, 1987
A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.
A:Reference number: A33135; MUID:87308013; PMID:3040685
A:Accession: B33135
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1475 <SHI>
A:Cross-references: UNIPROT:P09987; UNIPARC:UPI000014D972; GB:M17361; NID:g153639; PIDN:
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
submitted to the Protein Sequence Database, September 1990
A:Reference number: A33128
A:Accession: A33128
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-171,173-641,'N',643-1475 <SH2>
A:Cross-references: UNIPARC:UPI000017AC5E
A:Experimental source: strain GS-5
F:1036-1115/Domain: cpl repeat homology <CP1>
F:1224-1243/Domain: cpl repeat homology <CP2>
F:1289-1308/Domain: cpl repeat homology <CP3>
F:1354-1373/Domain: cpl repeat homology <CP4>
F:1419-1438/Domain: cpl repeat homology <CP5>
Query Match 54.9%; Score 4580; DB 2; Length 1475;
Best Local Similarity 56.4%; Pred. No. 1.2e-207;
Matches 896; Conservative 212; Mismatches 338; Indels 144; Gaps 18;
QY 1 MEKVRPFQWVKRKRWVTLVSASATMLASALGASVASADTDTA-----SDDSNQAVVTGQ 56
DB 1 MDKVRVYKLRVKRKRWTVSVASAVMTLTLLSGGLVKADSNESKSIQISNDSNTSVTANE 60
QY 57 TTN--NQATDQTSIAATAATSEQSASTDAATQASAAEQTQGTASTDTAAQTNNANEAK 114
DB 61 ESNVITEATSKQEAASSQTNHTVTTSSTSVVNPKE---VVSNPYTVGETASNGEKL- 115
QY 115 WVPTEENOGTDEMLAEA-----KNVATAESIIPSD-----LAKMSNVKQVDGKY 161
DB 116 -----NQQTTVTVTKTSEAAANNISKQTEADTVIDDSNANLQILKLPNVKSIDGKY 169
QY 162 YYYPDGNNKFNFAVSGDKIYYFDETGAYKDTSKVDADKSSSASVSQATIPAAANNRAYS 221
DB 170 YYDNNNGKVRNFTLIADGKLIHFDETGAYTDTSDTVNK--DIVTTSNLYKKYQVYD 227
QY 222 TSAXNFEAVDNYLTADSWYRPKSILKDGKTWTESGKODFRLLMAWVDPDTETFKRNVNYM 281
DB 228 RSAQSFEHVDHYLTAESWYRPKYILKDGKTWTQSTQKDFRPLMTWPDQETQRYVNYM 287
QY 282 NKVVGIDKTYTAETSOADLTAAELVQARIKQITSENNTKWLREASAFVKYTOQWNGE 341
DB 288 NAQLGINKTYDDTNQNLQNLNIAAATQAKIEAKITLKNKTDWLRQTIASFVKTSQSAWNSD 347
QY 342 SEKPYVDHLQNGALLFDNQDTLTPDTQSNYRLLNRTPTNQTGSLDSRFTYNPNDPLGGYD 401
DB 348 SEKPYVDHLQNGAVLYDNEGKLYTPVANSYRILNRTPTNQTGKKDPRYT--ADNTIGGYE 405

Qy 402 FLLANDVDSNPVQAEQLNLHLLNFGSIYANDADANFDSIRVDVANDVDADLLOISS 461
Db 406 FLLANDVDSNPVQAEQLNLHLLNFGSIYANDADANFDSIRVDVANDVDADLLOIAG 465
Qy 462 DYLKAAAGIGDKNNKANNHVSIVEASNDTPYLHDDGDNLMNMNKKFRLSMLSLAKPL 521
Db 466 DYLKAAAGIGHKNDKAANDHLSILEASNDTPYLHDDGDNLMNMNKKLRLSLFLAKPL 525
Qy 522 DKRSLNPLIHNLSVDREVEVTPSYSFARAHDSVODIIRDIKAEINPNISFGVS 581
Db 526 NQSGMNPILITNSLWRRDDNAETAAPVSFIRAHDSVODLIADIIRKAEINPNVGVIS 585
Qy 582 FTQBEIEQAFKIYNEDLKKTKKTHYNVPLSYTLNLTNGKSIPRVYIGDMFTDDGQYMA 641
Db 586 FTMBEIKKAFELYNKDLATEKCKTHYNTALSYALLTNKGSVPRVYIGDMFTDDGQYMA 645
Qy 642 NKTNYDAIESILKARKMYVSGGQAMQNYQNGEBILTSVRYGKALKQSKGDATRTTS 701
Db 646 HKTINYEATLLKARIKIVYSGGQAMRNOQVNSEIITSVRYGKALKATDTGDRTRTS 705
Qy 702 GVGVMGNQPNFSLDGK-VVALNMGAAHANOEYRALMVSTKDG VATYATDADASKAGLVK 760
Db 706 GVAVIEGNPRLKASRVRVVMGAAHKNQAYRPLLTITDNGIKAYHSDQEA--AGLVR 763
Qy 761 RTDENGILYFLNDDLKGVANPQVSGFLQVVPVPGAADQDRIVAASDTASTDGK--SLHQ 818
Db 764 YTNDRGELIFTAADIKGVANPQVSGFLGVVPVGAA--LIKMPALRLARPHQQMASVHQ 820
Qy 819 DAAMDSRVMPFGFSNFSQFATKEBEYTNVVIANNVDKFSVSGITDFEMAPQVYSTDQGF 878
Db 821 NAALDSRVMPFGFSNFSQFATKEBEYTNVVIANNVDKFAEWGVTDFEMAPQVYSTDGSF 880
Qy 879 LDSVIQNGYAFTRDYLQMSKANKYGTADQLVKAIKALHAKGLKVMADWVPDQWYTPKQ 938
Db 881 LDSVIQNGYAFTRDYLQMSKANKYGTADQLVKAIKALHAKGLKVMADWVPDQWYTPKQ 940
Qy 939 EVVTVTRTDKFKPIAGSQINHSYLVTTDTKSSGDDYQAKYGAFLDELKEXYPELFTKKQ 998
Db 941 EVVTVTRTDKFKPIAGSQINHSYLVTTDTKSSGDDYQAKYGAFLDELKEXYPELFTKKQ 1000
Qy 999 ISTQQAIDPSVKIKQWAKYNGNINILGRGADYVLSQVSNKYFNVA--STLFLPSLL 1056
Db 1001 ISTQQAIDPSVKIKQWAKYNGNINILGRGADYVLSQVSNKYFNVA--STLFLPSLL 1060
Qy 1057 GKVVESGIRVDGKVIYNSSATGDQVKASFITEAGNLYYFGKDGVMVTGAQTINGVYFF 1116
Db 1061 NQDSQVGFSDGKGVYVYST-SGYQAKNFTFISEGDKWYFNDNGYMTVGTGAQSGINGVYFF 1119
Qy 1117 LENGVALRNTIYTTDAQNSHYIYANDGKRYENGYQOF-GNDWRYFKDQGNMAYGLTVDGNV 1175
Db 1120 LSNGLQRLDALLKNEEDGTIAYGNDGRYENGYQFMGVSVRHFPANGEMSVGLTVIDGV 1179
Qy 1176 QYFDKGVQAKDKIIVTRDGKRVYFDQHNAAATNTFTIADKTHWYLGKDGVAVTGAQT 1235
Db 1180 QYFDEMGYQAKGVKFTVADGKIRYFDKQSGNMVRRNFIEBEGKWLVLGEDGAATVGSQT 1239
Qy 1236 VGRQKLYFEANGQOVKGFVTSDEGLKYFYDVSQDMWTDFTIEDKAGNWFYLGKDAV 1295
Db 1240 INGQHLVFRANGVQVKGFEVTDHHRISYIDGNSGDDQIRNRFVRAOQGVFYFNNGYAV 1299
Qy 1296 TGAOTIRGQKLYFKANGQVKGDIIVKGTGDKIRYVDAKSGBOVFNTKKAADGKTYVIGN 1355
Db 1300 TGARTINGQLLYFRANGVQVKGFEVTDHHRISYIDGNSGDDQIRNRFVRAOQGVFYF 1359
Qy 1356 DGVAVDPVSVKGVQTFKDSAGALRFPYNLKGVLVTCGSGWYETANHDWVYIQSGKALTG 1415
Db 1360 NGYAV-----TGARTI 1370
Qy 1416 NGQHLVFKEDGHQVKGQVLTGTDGKRVYDYDANSQDAFNKSVTVNGKTYTFGNDGTAQTA 1475
Db 1371 NGQHLVFRANGVQVKGFEVTDHHRISYIDGNSGDDQIRNRFV----- 1412

Qy 1476 GNPKGQTFKDGSDIRFYSGMEGLVTGSGWYENAOQWLYV-KNGKVLTLQTLQTVGSRVVF 1534
Db 1413 -----RNAQCGWFFYDNGNGYAVTGTARTINGHQLYF 1442
Qy 1535 DENGLOAKGKAVRTSDGKIRYFDENSQSMI 1564
Db 1443 RANGVQVKGFEVTDHHRISYIDGNSGGRV 1472
RESULT 3
JT0345
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)
N:Alternate names: sucrose 6-glucosyltransferase
C:Species: Streptococcus mutans
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 09-Jul-2004
C:Accession: JT0345; C31135
R:Ueda, S.; Shiroza, T.; Kuramitsu, H.K.
Gene 69, 101-109, 1988
A:Title: Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.
A:Reference number: JT0345; MUID:89137980; PMID:2976010
A:Accession: JT0345
A:Molecule type: DNA
A:Residues: 1-1375 <UED>
A:Cross-references: UNIPROT:P13470; UNIPARC:UPI0000155515
A:Experimental source: GS-5
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
J. Bacteriol. 169, 4263-4270, 1987
A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.
A:Reference number: A33135; MUID:87308013; PMID:3040685
A:Accession: C33135
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-349 <SHI>
A:Cross-references: UNIPARC:UPI000014E25D; GB:M17361
C:Genetics:
A:Gene: gtfC
C:Function:
A:Description: catalyzes the synthesis of both water-soluble and water-insoluble glucans
C:Keywords: duplication; glycosyltransferase; hexosyltransferase
F:1-34/domain: signal sequence #status predicted <SIG>
F:35-1375/product: Glucosyltransferase #status predicted <MAT>
F:1126-1145/domain: cpl repeat homology <CP1>
F:1253-1272/domain: cpl repeat homology <CP2>
F:1318-1337/domain: cpl repeat homology <CP3>
Query Match 50.5%; Score 4214.5; DB 2; Length 1375;
Best Local Similarity 60.3%; Pred. No. 1.7e-190;
Matches 811; Conservative 186; Mismatches 296; Indels 53; Gaps 14;
Qy 1 MEKNVRFKMHKKRWTLSVASATMLASGASVASADTDTASDSSNQAVVTGDT--T 58
Db 1 MEKVRFLRKVKRWTVSASAVTLTSLGSLVAD---STDDRQAVTESQASLVT 57
Qy 59 NNOATDTSIA---ATATSEQASDAATDOASAEQQTGTTAST-DTAAQTTTNNANBA- 113
Db 58 TSEAAKETLTATDSTATSQPTATVTDNVSTTNQSTNTTANTANFVVPVPTTSEQAK 117
Qy 114 -----KWVPTENEN---QGFTDEM-----LABAKNVATAESDS 143
Db 118 TDNSDKIITTSKAVNRLTATGKFPANNNTAHPKTVTDKIVPIKPIKGLKQPSLSQDD 177
Qy 144 IPSDLAKMNVKQVGDGKYVYDQGNVKQNFASVGDKIYVFDGTGAYKDTSKVDADKSS 203
Db 178 IAA-LGNVKIRKNGKYYKEDGTQKNTALNKGKTFEFTDEGALSNNT-LPSKGN 235
Qy 204 SAVSQNATIFAANNRAYSTSAKNFPAVDNVLTAWSYRPSILKDGKTTWESGDDFRPL 263
Db 236 ITNNDNTNSPAQYNQVSTDVANFEHVDHYLTASWYRPSYILKDGKTTWSTQTEKFRPL 295
Qy 264 LMAWVPTETKRNVTNPNKVVIGIDKTYTATSTQADLTAAABLQVARIQKITSNNYTKW 323
Db 296 LMTWVPDQETQRQVYVYNNALQGIHQTYNTATSPQLNLAAQTIQTKIEBKITAENKTNW 355

```
QY 324 LREAIASFVKTPQWNGESEKPYDDHLQNGALLFDNQDTDLTPDTCNSYRLNRTPTNQTG 383
Db 356 LRQTSIAFVKYQSAWNSSEKPFDDHLQKGALLYSNNSKLTQSAWNSYRLNRTPTNQTG 415
QY 384 SLDRFTYNNPNDPLGGYDFLLANDVDNSNPVQAEQLNWLHLYNFGSIYANDADANFDS 443
Db 416 KKDPRT--ADRTIGGYEFFLLANDVDNSNPVQAEQLNWLHLYNFGSIYANDADANFDS 473
QY 444 IRVDAVNDVADLLQIISDDYLKAAAGYIDKNKNANNNHVSIVEAWSNDNTPYLHDDGDNL 503
Db 474 IRVDAVNDVADLLQIAGDYKAAAGIHKNDKAAANDHLSILEAWSNDNTPYLHDDGDNMI 533
QY 504 NMDNKFRLSMLWSLAKPLDKSGNLPLIHNLSIVREVDREVEVTPSVSPARAHDSVQD 563
Db 534 NMDNRLSLLSYLAKPLNQRSGMNPPLITNSLVNRDNDNAETAAPVPSYFTRAHDSVQD 593
QY 564 IIRDIKAEINPNFGYSFTOEIEQAPKIYNEDLKTKDKKYTHVNPVLSYTLTLTNKGS 623
Db 594 LIRNIIRTEINPNVGVSYFTTEEIKKAEIYNKOLLATEKKYTHYNTALSYALLTNKSS 653
QY 624 IPRVYGDMTDDGOYMAKNTVNYDAIESLLKARKMYVSGQAMQNYQIGNGEILTSVRY 683
Db 654 VPRVYGDMTDDGOYMAHKTINYEAIETLLKARIKYVSGQAMRNQGVGNSEIITSVRY 713
QY 684 KGKALQSDKGDATRTTSGVGVWGNQPNFSLDGK-VVALNMGAAHNOEYRALMVSTKD 742
Db 714 KGKALKATDTGDRTRTSGVAVIEGNNPSRLKASDRVVVNMGAHKNQAYRPLLLTTDN 773
QY 743 GVATYATADAKSKAGLVKRTDENGVLVFLNDDLGKVANPQVSGFLQVWVPVGAADDQDIR 802
Db 774 GIKAYHSQDEA--AGLVRYTNDRGELIETAADIKGYANPQVSGVLGVWVPVGAADDQVR 831
QY 803 VAASDTASTDGKSLHODAMDSRVNFEFGSNFQSPATKEEBYTNVVIANNVDKVFWSGIT 862
Db 832 VAASAPSTDGKSHONAAALDSRVNFEFGSNFQSPATKEEBYTNVVIANNVDKVFWSGIT 891
QY 863 DFEMAPQVYSTDGOFLDSVTLQNGYAFTRDVLGMSKANKYGTADOLYKAIKALHAKGLK 922
Db 892 DFEMAPQVYSTDGSFLDSVTLQNGYAFTRDVLGMSKANKYGTADOLYKAIKALHAKGLK 951
QY 923 VMADWVPQMYTFPKQEVVTVTRTDKFGKTAGSQINHSLSVTVDTKSSGGDDYQAKYGGAF 982
Db 952 VMADWVPQMYALPEKEVTVATRVDTKYGTPVAGSQIKNTLVVDGKSGKQQAKYGGAF 1011
QY 983 LDELKKEYPELFTKQISTGQAIDPSVKIKQWSAKYFNGSNILRGADYVLSQDVSNKYF 1042
Db 1012 LEELQAKYPELFAKQISTGVPMDSVKIKQWSAKYFNGSNILRGAGYVLDQQAINTYF 1071
QY 1043 NVASDTLPLPSLLG----KVVESGIRYDGKGIYNSSATGDQVKASPFITEAGNLYYFG 1097
Db 1072 SLVSDNTFLPKSLVNPNNHGTSSSVTGLVFDGKGYVYVYST-SGNQAKNAFISLGNWNYFD 1130
QY 1098 KDGWVYVYCAQTINGANYFFLENGTALRTIYTDAGNSHYVANDGKRYENGQYQFGNDWR 1157
Db 1131 NGWVYVYCAQTINGANYFFLENGIOLRNAIYDNGKVLVSYGNDGRRYENGYYLFGQWR 1190
QY 1158 YPKDGNMAVGLTVDGNYQYFDKGVQAKKIIVTRDGKVRIFYDOHNGNAATNTFIADKT 1217
Db 1191 YFQNGIMAVGLTRVHGAVOYFDASGFQAKGQFITTADGKLRYFDRDSGNQISNRFVRNSK 1250
QY 1218 GHVYVLKDGVAVTGAQTVGQKLYFEANGQVKGDFVTSDEGKLYFYDVGSDGMWTF 1277
Db 1251 GEWFLFDHNGVAVTGTVTFNGORLYFKPNGVQAKGEFIRDANGYLRYVYDPSNGNEVRNF 1310
QY 1278 IEDKAGNWFYLGKGAAYTGAQTIRG 1303
Db 1311 VRNSKGWFLFDHNGIATGARVNG 1336
```

RESULT 4
A45866
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans
C,Species: Streptococcus mutans

C,Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Mar-2004
C,Accession: A45866
R,Honda, O.; Kato, C.; Kuramitsu, H.K.
J. Gen. Microbiol. 136, 2099-2105, 1990
A,Title: Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the glucosyl
A,Reference number: A45866; MUID:91100958; PMID:2148600
A,Accession: A45866
A,Status: preliminary
A,Molecule type: DNA
A,Residues: 1-1431 <HON>
A,Cross-references: UNIPARC:U0100017AC5C; GB:M29296
C,Keywords: glycosyltransferase; hexosyltransferase
P,181-201/Domain: cpl repeat homology <CP1>
P,1127-1146/Domain: cpl repeat homology <CP2>
P,1192-1211/Domain: cpl repeat homology <CP3>
P,1257-1276/Domain: cpl repeat homology <CP4>
P,1277-1297/Domain: cpl repeat homology <CP5>
P,1321-1340/Domain: cpl repeat homology <CP8>
P,1341-1361/Domain: cpl repeat homology <CP6>
P,1385-1404/Domain: cpl repeat homology <CP7>

Query Match 43.1%; Score 3599; DB 2; Length 1431;
Best Local Similarity 48.4%; Pred. No. 1.7e-161;
Matches 724; Conservative 241; Mismatches 414; Indels 118; Gaps 26;
QY 1 MEKVRPMHKVKRWVTLVSASA--TWLSALGASVAS-----ADTDTS 44
Db 1 METKRYRMHKVKRWVTVAVASGLITLTGTTLSSVSAAETEQTSDKVVTKSEDDKAA 60

QY 45 DDSNOA--VVTGDTNNQATDTSIAATATS--EQSASTDAATDQASAAEQTCGTTAST 100
Db 61 SSSQTPAKTKQAQTQTQAOQSOANVADTSTITKETPSQNIITQANSDDKTVNTKSE 120
QY 101 D--TAAQTNTNANBAKVPPTENENQGTDEMLBAKNVAT--AESDISPSDLAKMSNVKQV 157
Db 121 EAQTSSEERTKQSEBAQ--TTASSQALTQAKAELTKQRTAAQENKGNPVDLAAIPNVKQI 177
QY 158 DGKYYVYDQDGNVKNFAVSGDKIYVDE--TGAYKOTSKVDADKSSSAVSQNAITFAAN 216
Db 178 DGKYYVYDQDGNVKNFAVSGDKIYVDE--TGAYKOTSKVDADKSSSAVSQNAITFAAN 234
QY 217 NRAYSTAKNFEADVNTLTDADSWRPKSILKDGKTWTESGKDDPRLLMAWMPDTEKRN 276
Db 235 NQIVNFPENTSLTIDNVTADSWRPRKDLKNGKWTATASSEDRLPLMSWHPKQTOIA 294
QY 277 YVNTMNVK-VGIDKTYTAETSQADLTAAAEILVQARIQKITSNNNTKWLREISAISFVKQ 335
Db 295 YLNTMNVQGLGTGENYTADSSQESLNLAQAQTVQVKIETKISQTCQTLWRLDIINSFVKQ 354
QY 336 PQWNGESEKPYD----DHLQNGALLFDNQDTDLTPDTCNSYRLNRTPTNQTGSLDSRFTY 391
Db 355 PNWNSQTESDTSAGEKHLOGGALLYSN-SDKTAYANSDYELLNRTPTTSQTG----KPKY 409
QY 392 NPNDPLGCGYDFLLANDVDNSNPVQAEQLNWLHLYNFGSIYANDADANFDSIRYADV 451
Db 410 PEDSSGCGYDFLLANDVDNSNPVQAEQLNWLHLYNFGSIYANDPEANFDCVRVADV 469
QY 452 VDADLLQISSDYLKAAAGYIDKNKNANNNHVSIVEAWSNDNTPYLHDDGDNLNMDNKFRL 511
Db 470 VNADLLQIASDYLKAAAGYIDKNKNANNNHVSIVEAWSNDNTPYLHDDGDNLNMDNKFRL 529
QY 512 SMLWSLAKPLDK-----RSGNLPLIHNLSIVREVDREVEVTPSVSPARAHDSVQD 563
Db 530 SLLYALTRPEKDNASKNNEIRSGLEPVITNSLNRSRASEGKNSEMANVYIFTRAHDSVQ 589
QY 564 IIRDIKAEINPNFGYSFTOEIEQAPKIYNEDLKTKDKKYTHVNPVLSYTLTLTNKGS 623
Db 590 VIAKIIRKQINPKTDGLTFTLDELKQAFKIYNEDMRQAKKYQTSNIITAYALMSNKS 649
QY 624 IPRVYGDMTDDGOYMAKNTVNYDAIESLLKARKMYVSGQAMQ-NYQIGNGE----- 676
Db 650 ITRLYYGDWYDSDGOYMATKSPYYDAIDTLKARIKYAAGQDMKITVYEGDKSHMDWDY 709

```
QY 677 ---ILTSVRYGKALQKSDGDAATRTTSGVGVVGNQPNFSLD-GKVALNMGAHANQBY 733
Db 710 TGVLTSVRYGTGANETKQSEATKTQGMVITSNPNSLKNQNDKVIYMMGAHKNQBY 769
QY 734 RALMVSTKDGVAATADASAGLVKRTDENGILYFLNDDLKGVANPOVSGFLOWVPV 793
Db 770 RELITTKDGLTSYSDAAK--SLYRKTNDRGELVFDASDIOGYLNPQVSGVLAVWVPV 827
QY 794 GAADQDQIRVAASDTASTDGKSLHDAAMDSTRVMEFGFSNFQSFATKBEYETNVVIANV 853
Db 828 GASDNQDVRVAASNKANATQVYESSALDSQLIIEGFSNFQDFVTKDSYTNKKIAQNV 887
QY 854 KXPVSGITDFEMAPQYVSTSGQFLDSVIONGYAFTDRYDLGMSKANKYGTADOLVKA 913
Db 888 QLFKSGVTSFEMAPQYVSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSQQDMINAV 947
QY 914 KALHAKGLKVMADWPDQMYTPKGEVTVTTRDKEGPIAGSQINHSLVYVTDKSSGD 973
Db 948 KALHSGGIQVIADWPDQYINLPGEVTVTATRVNDYGEYKDSSEIKNTLYAAANTSKNGD 1007
QY 974 YOAKYGGAFDLDELKELPELTKQISTQQAIDPSPVKIQMSAKYFNGSNILGRGADVYL 1033
Db 1008 YOAKYGGAFSLAAYKYSIFNRTQISNGKIDPSEKITAWAKYFNGNINILGRGVYVL 1067
QY 1034 SDQVSNKYFNVASDTLFLPSSLLGKVESGIRYDKGYIYNSSATGDOQVKASFITEA-GN 1092
Db 1068 KDNASDKYFELKGNQYLPKQMTNKEASTGFVNDGNGMTFYST-SGYQAKNSFVQDAKN 1126
QY 1093 LYFEGKDGVMYGAQTINGANYFPLENGTALNTIYTDAGNSHYVANDGKRYENGYOQF 1152
Db 1127 WYFDPNNGHMYGLQOLNGEVQYFLSNGVQLRESFLENADGSKNYFGLHGNRYSNGYYSF 1186
QY 1153 GND--WRYF-KDGNMAGLTLTVDGNVQYFDKDGVAQKDKIIVTRDGKVRYPDOHNGNAAT 1209
Db 1187 DNDKWRVFDASGVMAVGLKTINGTQYFDQGYQVKGAWITGSGKRYFDDGSGNNAV 1246
QY 1210 NTFIADTKGHYYLKGQVAVTGAQTVGKQKLYFEANGQVQKGFVTSDEGKLYFYDVDS 1269
Db 1247 NRFANDKNGDWYLNLSGDIALVGQVQTINGKTYFYQDGQKQIKGKIIT-DNGKLKYFLANS 1305
QY 1270 GDMWTDTEBQKAGNWFYLGKGAQVGAQTIRGOKLYFKANGQVQKGDIVKGTGKIRY 1329
Db 1306 GELARNIFATDSQNNWYFYGSDGAVTGSQTIAGKKLYFASDGQVQKGSFVT-YNGKVHY 1364
QY 1330 YDAKSGEVOFNKVAADGKTYVIGNDGVAVDPVSVVKGQTFKDGASGLRFFYNLQKQVLTG 1389
Db 1365 YHDSGELQVNRFEADKDG----- 1383
QY 1390 SGWYETANHDWYIOS-KGALTEQTINGOHLYPKEDGHQVKGQVLTGTDGKVRYYD 1445
Db 1384 -----NWYILDSNGEALTGSRINDQRVFFFTREGKQVKGDVAYDERGLLRYD 1431

RESULT 5
A44811
A;Glucosyltransferase (EC 2.4.1.-) I - Streptococcus salivarius
C;Species: Streptococcus salivarius
C;Date: 31-Dec-1993 #sequence revision 31-Dec-1993 #text_change 09-Jul-2004
C;Accession: A44811; S22726; S28809
R;Giffard, P.M.; Simpson, C.B.; Milward, C.P.; Jacques, N.A.
J. Gen. Microbiol. 137, 2577-2593, 1991
A;Title: Molecular characterization of a cluster of at least two glucosyltransferase genes
A;Reference number: A44811; MUID:92148377; PMID:1838391
A;Accession: A44811
A;Molecule type: DNA
A;Residues: 1-1518 <GIF>
A;Cross-references: UNIPROT:Q00600; UNIPARC:UPI00000BEF31; EMBL:Z11873; NID:G47526; PIDN
A;Note: sequence extracted from NCBI backbone (NCBI:81050, NCBI:81052)
C;Genetics:
A;Gene: gtfu
C;Keywords: glycosyltransferase; hexosyltransferase
F;1307-1326/Domain: cpl repeat homology <CP4>
```

```
Query Match 39.8%; Score 3325.5; DB 2; Length 1518;
Best Local Similarity 45.0%; Pred. No. 1.4e-148;
Matches 710; Conservative 239; Mismatches 456; Indels 173; Gaps 33;

QY 1 MEKNVRFKMHKKWVTLVASATMLASALG-----ASVASAD-----TDTASDD---- 46
Db 1 MENKTHYKHLKHKWMTVIASVA-LATVUGGLSVTTSSVSADSETQDKTIVTQSNSGTTA 59
QY 47 ----SNQAVVTGDTQTNNOATD-----QTSIAATAT-----SQBSASTDAATDOASAA 90
Db 60 SLVTSPEATKADKKTNTKEADVLTPAKETNAVEATTNTQATAEATATTATADAVAA 119
QY 91 ETOQTGTTASTDTAAQTNNANE-AKWVPTENENOGFTDEMLAEAKNVATAESDPSIDIA 149
Db 120 VNKAEAVTTDAPAVTEKAEQPAVKAEVNV-----TEVKAPEAALKSEVEAAL 171
QY 150 KMSNVKQVGDGKYVYDQDGNVKNKFAVSGDKIYYPDETGAYKDTSKVDADKSSSAVSQN 209
Db 172 SLKNTIKNDGKYVYVNEDEGSHKENPAITVNGQLLYFGDKGALTSSSTYSFTTGTNNIVDG 231
QY 210 ATIPAAKNRAYSTSAKNFEADVNYLTADSWRPKSILKDGKTWTSKGDDPRPLMAWMP 269
Db 232 ---FSINRAVDSSEASFELIDGYLTADSWRPAIIKDGVTWQASTAEDFRPLMAWMP 288
QY 270 DTETKRNVTYNNKVVGIDKTYTATSTQADLTAAAEVQARIEQKITSENNTKWLREATS 329
Db 289 NVDTQVNYLNTMSKVFNLDKAYSTTDKQETLKVAAKDIQIKIEQIKQAEKSTQWLRETS 348
QY 330 AFVKTQPOWNGESEKPY-----DHLQNGALLFDNQTLTPDTQSNRYLLNRTPTNQTS 384
Db 349 AFVKTQPOWNETEN-YSGKGEGDHLOGGALLYVND-S-RTPWANSDYRLNRTATNQGT 406
QY 385 LDSRFTYNPNDP--LGGYVDFLLANDVNSNPVQAEQLNWLHLLNFGSIYANDADANFD 442
Db 407 IDKSILDSQSDPNMGGFDFLLANDVLSNPVQAEQLNQHLYLNMWGSIVMGDKDANFD 466
QY 443 SIRDAVDNVDADLLQISSDYLKAAYGIDKNNKNANHVSIIVEASDNDTPYLHDDGDNL 502
Db 467 GIRDAVDNVDADMLQLYNTYFREYGVNKGSEANALAHISVLEAWSLNDNHNDKTDGAA 526
QY 503 MNMDNKFSLMLWSLAKPLDKRS-GLNPLIHNSLVDRVDDR-----EVETV 548
Db 527 LAMENKQRLALLFLAKPIKERTPAVSPLYNNTFTTQORDEKTDWINKDGSKAYNEDGTV 586
QY 549 P-----SYSEFARAHDSVQDIIRDIKAEINPNPSFGYSFTOEETEQAFK 592
Db 597 KQSTIGKNEYKGASGNVTFIRAHNNVQDIIAEIIKKEINPKSDGFTITDAENKQAF 646
QY 593 IYNEDLKKTKKTYHNVPLSYTLNLTNKGSIPIRVYGDMEFTDDGQYMANKTVNYDAIES 652
Db 647 IYNKMLSSDKKTYLNNIPAAVAVMLQNWETITRVYGDLYTDDGHYMETKSPYDTIVN 706
QY 653 LLKARKMFKVSGQAMQNYQI-----NGEILTSVRYGKALQKSD-KGDAT 697
Db 707 LMSRIKYVSGQQAQSYWLPDTGKMDSNDVELYRTNEVYTSVRYGKQIMTANDTEGSKY 766
QY 698 TRTSQGVVMGNQNFSLDQKV-VALLNMGAAHNOEVALMVSTKDVATYATADASKA 756
Db 767 SRTSQGVTVANNPKLNDQSAKUNVENMGKIHANQYRALIVGTADGKNFTSDADAIAA 826
QY 757 GLVXRTDENGILYFLNDDLKGVANPOVSGFLQVWVPVGAADDQDQIRVAASDTASTDGK-S 815
Db 827 GYVKTDSNGVLTGANDIKGEYFDMGSGFVAVVWVPVGCASNDQDQIRVAPSTEAKKEGELT 886
QY 816 LHQDAAMDSTRVMEFGFSNFQSF--ATKBEETNVVIANVNDKVFVSWGITDEMAPQYVSS 873
Db 887 LKATEAVDSQLIIEGFSNFQTIIPDGSOPSVYTNKRAENVLDLFXKSWGVTSEMAPQFVSA 946
QY 874 TDGQFLOSVIONGYAFTDRYDLGMSKANKYGTADOLVKAIKALHAKGLKVMADWVPDQMY 933
Db 947 DDGTFLOSVIONGYAFADRYDLAMSNNKYSGKEDRLDALKHAKGIQATADWVPDQIY 1006
QY 934 TFPKQEVVTVTRTDKFGKPIAGSQINHSLYVYTDTKSSGDDYOAKYGGAFDLDELKELPEL 993
```


Db	1007	QLPGKEVTAETDAGRKIAADIADHSLYVANSKSGKDQAKYGGFELAEKAKYPM	1066
Qy	994	FTKKQISTQQAIDPQVKIKQMSAKYFNQSNILGRGADYVLSQVSNKYFNVASDTFLPLPS	1053
Db	1067	FKVNMISTGKPIDDSVKLQKMAEYFNCTNVLERGCVGLSDEATGKFTVTKEGNFIPL	1126
Qy	1054	SILGK-VVESGIRYDGKGIYNNSSATGQVKASFIETAGNIYYFKGQYMTGAQ-TING	1111
Db	1127	QLTGKEKVTGSSDGKITY-FGTSGTQAQSAFVTFNGNTYFDPARGHMVTNSEYSPNG	1185
Qy	1112	AN-YPFLENGTALRNTIYTDAGNSHYANDGKRYENGQQF-----GND-----WR	1157
Db	1186	KDVRFLPNGIMLSNAFYIDANGNTYLYNSKGWYKGGYTFEDVSETDKDGKESKVVKFR	1245
Qy	1158	YP-KDGNMAVGLTTVDGNVQVPDKGVQAKOKIIVTRDGKVRYPFOHNGNAATNTFIADK	1216
Db	1246	YFTNEGVMAGVTVTDGFTQYFGEDGFQAKOK-LVTFKGTYYFDPHTGNGIKDWTW-RNI	1303
Qy	1217	TCHWYVLKDGVAVTGAQTGKQKLYPEANGQOVKGDFTSDGKLYFYDVSQDMWTD	1276
Db	1304	NGKWTYFDANGVAATGAOVINGQKLYFNEGDSQVKGVVKNADGTYSKYKGFGLVTNE	1363
Qy	1277	FIEDKAGNWFYLGKDGAAVTGAQTIRGQKLYFKANGQOVKGDIVKGTGDKIRYIDAKSGE	1336
Db	1364	PFTTDGNVWYAGANGKTVTGAOVINGQHLVFNADGSQVKGGVKNADGTYSKYNASTGE	1423
Qy	1337	QVFNKTVKAADGKTVYVINGDVAVDPSPVVGQTGTFKDGASALRFYNLKGQLVGSGWYETA	1396
Db	1424	RLTNE-----FETG	1433
Qy	1397	NHDWYVI-QSGKALTGEQTINGQHLVFKEDGHQVKGQLVGTGDKVRYYDANSQDAENK	1455
Db	1434	DNNWYIIGANGKSVTGEVKIGDITVFFAKQGVKQVGTVSAGNGRISYIYGDGSKRAVST	1493
Qy	1456	SVTVNGKTY-YFGNDGTA	1472
Db	1494	WIEIQPGVYVFDKXGLA	1511
RESULT 6			
T30858			
Glucosyltransferase - Streptococcus salivarius			
C:Species: Streptococcus salivarius			
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004			
C:Accession: T30858			
R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.			
Infect. Immun. 63, 609-621, 1995			
A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for pri			
A:Reference number: 220909; PMID:95122197; PMID:7822030			
A:Accession: T30858			
A:Status: preliminary; translated from GB/EMBL/DBJ			
A:Molecule type: DNA			
A:Residues: 1-1577 <SIM>			
A:Cross-references: UNIPROT:Q55265; UNIPARC:UPI00000B08087; EMBL:L35928; NID:G662380; PID			
A:Gene: gtfm			
Query Match 38.9%; Score 3245; DB 2; Length 1577;			
Best Local Similarity 42.0%; Pred. No. 9.4e-145;			
Matches 714; Conservative 236; Mismatches 486; Indels 264; Gaps 34;			
Qy	1	MEKNVRPKQHKVKKRWTLVSASATMLASALGASVA-----SADTDTA-----SDS	47
Db	1	MENKVRPKLHKYKKNWVTIGVTTLSMVALAGSLLAQGKVEADETSAPNGDLQQLSBDG	60
Qy	48	NOAVVTGDTNNQATDQTSAAATSEQS-----ASTDAATDQASAAEOTQGTASTD	101
Db	61	TASLTV-TTITTEQASQASASAVATASVSUHSHTSFQAATSAVSQATQAQT-SPVASQE	118
Qy	102	TAAQTNNANAKWVPTENENQGTDEMLA-----EAKNVATA	139
Db	119	VAVSSQTQSSGQETQTTEQVSSQGTSTQVAGTSAQTSAPSVTEQARPRVLTNAAPAIATR	178

Qy	140	ESDSI-----	144
Db	179	AADSTIRINARNRNTIITASGTTPNVTIITGNTPKPNVTVTSNNGTRPNVTIITQPNQ	238
Qy	145	-----PSDLAK-----MSNVQVQDGKYYYYDQDGNVKKNFVAVSGDK	181
Db	239	PNKVPQSPQPNKVPQPNQPSLDYKPVASNLKTIIDGQYV-ENGVVKKNAATELDR	297
Qy	182	IYYFDETGAYKDTGK-----VDADKSSSSAVSQNATIPAAANNRAYSTSAKNEFAVDNYLAD	237
Db	298	IYYFDETGAMVDQSKPLRAD-----IPNNSIYAVYNOAYDTSKSFHLDNFLAD	350
Qy	238	SWYRPSKTLKDGKWTBESGKDDFPLLMAMWPPDTETKENYUNYNNKVVIGIDKTYTAETISQ	297
Db	351	SWYRPKQLLKDGKWTASTEDKYRPLMTWMPDKVTQVNYLYNMSQQGFKNTYTTDMS	410
Qy	298	ADLTAALAEVQARIEQKITSENNTKWLREALSAFVKTPQPNWNGSEKPY--DDHLONGA	354
Db	411	YDLAAAEVTVORGIEERIGREGNTTWLQMLMSDFIKTPGWNSEEDNLLVKDHLQGA	470
Qy	355	LLFQNTDLTDPDTOSVRLNRTPTNQTGSLDSRFTYNNPNDPLGQYDFLLANDVDSNPV	414
Db	471	LTFLNNS-ATSHANSDFRLMNRTPNTQTGTR---KYHIDRSNGGYELLANDIDNSNP	525
Qy	415	VOAQNLWLHYLLNFGSIYANDADNPSIRVDADVNDVADLLQIISDYLKAAYGIDKNN	474
Db	526	VOAQNLWLHYIMNIGSILGNDPSANFDGVRIDAVNDVADLLQIASDYFKEKYKVAONE	585
Qy	475	KNANNHVSIVBEASDNDTPTYLHDDGDNLMMNDKFRLSMLSLAKPLDKRSGNLPHNS	534
Db	586	ANATAHLSILEAWSYNDHYNKDTKGAQLSDNPLRETLTTFLRKSNYRGLSRLVINS	645
Qy	535	LVDRVDDREVEVTPSYSPARAHSEVODIIRDIKABINPNSFGYSTQBEIQAQFIY	594
Db	646	LNNRSSEKHTPRDANYIFVRAHDSVQAVLANIISKQINPKDTGFTTMDLQKAFEY	705
Qy	595	NEDLKKTDKKTYHYNVPLSYLLTNKGSIRVYVYGMFTDDGQYMANKTVNYDAIESLL	654
Db	706	NADIAKADKKTYQNI PAAYATMLTNKDSITRVYVYGLFTDDGQYMAESKSPYNAIDALL	765
Qy	655	KARMKYVSGQAMQNYQIGNGEILTSVRYKGALKQSDKGDATRTSGVGMGNQPNFS	714
Db	766	PARIKYVAGQDMKVTKNGYEIMSSVRYKGAEANQGTAEENQGLVLTANRPDMK	825
Qy	715	LDGK-VVALNMGAAHQAQYRPMVSTKDGCVATYATDADASKAGLVKRTDENGYYFLND	773
Db	826	LGANDRLVVMNGAAHQAQYRPMVSTKDGCVATYATDADASKAGLVKRTDENGYYFLND	884
Qy	774	DLKGVANPQVSGFLOVWPVCAADDODIRVAASDTASTDGKSLHQDAAMDSDRMEGFSN	833
Db	885	DIAGHSTVEVSGYLAVWPVPGASENQDARTKASSTKGE-QVFESSALDSQVIYEGFSN	943
Qy	834	FQSPATKEEYTNVIVANNVDPFVSWGITDPEMAPQYVSSSTDGQFLDSVIQNGYAFTRY	893
Db	944	FQDFVKTPSQYTNVIAQNAKLFKEWGITSEFAPQYVSSQDGTFLDSIIEGYAFEDRY	1003
Qy	894	DLGMSKANKYGTADQLVKAIKALHAKGLKVMADWVPDQMYTFPQKQVTVTRTKFGKPI	953
Db	1004	DIAMSKNKYKSLKDLMDALRALHAEGISATADWVPDQIYNLPGKEVTVTSRNTSYGTPR	1063
Qy	954	AGSOINHSLYVDTKSSGDDYQAKYGGAFDLDELKKEYPELTKKQISTGQALDPSVKIKQ	1013
Db	1064	PNABEINSLYAAKTRTFGNDPQGGYGAFLDELKAKYPAIFERVQISNGRKLTTNEKITQ	1123
Qy	1014	WSAKYFNGSNILRGADYVLSQVSNKYFNVASDTLFLPSSLLGKVSVSGIRYDGKGIY	1073
Db	1124	WSAKYFNGSNILRGADYVLSQVSNKYFNVASDTLFLPSSLLGKVSVSGIRYDGKGIY	1182
Qy	1074	NSSATGQVQKASFTTEAGN-LIYFGKQGYMTGAQTINGANYFFLENGTALRNTIYTDAQ	1132
Db	1183	-LSIGGVYLAQNTFIQVGANOWYFDPKNGNMVTVGQVIGDKKYFFLDNGLQRLHVRQSGD	1241

```

QY 1133 GSHYYANDCKRYENGYQQFG---NDWRYPKQGN--MAYGLTTVDGNVQYFD-KGQVQAK 1186
DB 1242 GHVYYDPRKQVAFNGFYDFAGPRQVRYPF-DGNGQMYRGLHDMYGTTFYFDEKTIQAK 1300
QY 1187 DKLIIVTRDGRVRYFDQHNGNAANTFIAD-KTGHWTYLGKQGVAVTGAOTVKGQKLYFEA 1245
DB 1301 DKFIRPADGRTRIFPDITGNLAVRFAQNPENKAWYILDSNGAVTGLQTINGKQYFDN 1360
QY 1246 NGQOVKGFVTSDEGKLYFYDVSDDMTDTFIEDKAGNWFYLGKDGAAVTAQTIRGOK 1305
DB 1361 EGRQVKGFHFTVNNQR-YFLDGSGBIAPSRFVTEN-NKYYVYDVGKLVKGAQVINGNH 1418
QY 1306 LYFKANGQOVKGDIVKGTGKTRYYDAKSGQVFNKTVKAAADGKTVVIGNDGVAVDPSV 1365
DB 1419 YFNNNDYSQVKGAWANG-----RYDGDGSGQAVSNQFI-----1451
QY 1366 KGQTFKADSGALRFYNLKLQVLGTGSGWYETANHDWYI-QSGKALTGEOTINGOHLYPE 1424
DB 1452 -----QIANQWAYLNDQGHKVTGLQNNKVVYFGS 1483
QY 1425 DGHQVKQLVTGDKRVRYDANSQDQAFNKSVTYNGKTYFNGDGTAGNPKGQTEK 1484
DB 1484 NGAQVKGKLLT-VQGKKCYFDAHTGQVVRV-----1515
QY 1485 DGSDFRYSMEGQLVTGSGWYENAOQWLYVKN-GKVLTLGLQTVGSGQRYVFDENGIOAKG 1543
DB 1516 -----EARGCWYTFNSAGQAVTQQVINGKQLYFDGSGRQVKG 1554
QY 1544 KAVRTSDGKIRYFDENSGSM 1563
DB 1555 RYVYVG-GRKLFCDKAKTGSL 1573

RESULT 7
S22737
Glucosyltransferase (EC 2.4.1.-) S - Streptococcus salivarius
C:Species: Streptococcus salivarius
C:Date: 16-Sep-1992 #sequence revision 16-Sep-1992 #text_change 09-Jul-2004
C:Accession: S22737; S28810; B44811; S22727
R:Jacques, N.
submitted to the EMBL Data Library, March 1992
A:Reference number: S22726
A:Accession: S22737
A:Molecule type: DNA
A:Residues: 1-1599 <JAC>
A:Cross-references: UNIPROT:Q00599; UNIPARC:UPI00000BEF34; EMBL:Z11872; NID:g47530; PIDN
A:Experimental source: ATCC 25975
R:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.
J. Gen. Microbiol. 137, 2577-2593, 1991
A:Title: Molecular characterization of a cluster of at least two glucosyltransferase genes
A:Reference number: A44811; MUID:92148377; PMID:1838391
A:Accession: S28810
A:Molecule type: DNA
A:Residues: 1-51 <GIF>
A:Cross-references: UNIPARC:UPI000017027C; EMBL:Z11873
C:Genetics:
A:Gene: gtfK
C:Keywords: glycosyltransferase; hexosyltransferase
F:1456-1475/Domain, cpl repeat homology <CPR>

Query Match 38.5%; Score 3214.5; DB 2; Length 1599;
Best Local Similarity 41.3%; Pred. No. 2.6e-143;
Matches 699; Conservative 267; Mismatches 516; Indels 211; Gaps 41;

QY 1 MEKNVRFKHKVKKWVLSVASATWLASALGASVA-----SADTDTASD 45
DB 1 MENKRYKLHKVKKQWVTLAVASVA-LATIVGSGVATSLASAEETNNSGSPSTTVGE 59
QY 46 DSNQAV-----VTGD--OTTNNQATDQTSIAAT-----ATSEQASASTDAATDQ 86
DB 60 NTNPFVEKEVGTTTETVANTTSATTTTERRAEVTKADKAGITTVQNPNSGTTSDRAAAVEAKP 119
QY 87 ASAAEQTQGTASTDTAAQTNTTANAEKWPPTENENQGFDEMLAEAKNVATAESDSIPS 146

```

```

DB 120 ETTAKPEVATKPEATATTSEVAANAGVA---APTTEKS-----KELSEAEIKAAVSLDNIK- 171
QY 147 DLAKSNVQKQVKKYVYDQDGNVKNFPAVSVDKIYYFDETCAYKDTKVDKADKSSAV 206
DB 172 -----KEKDGKYYLLEDSHKKNFALTIVNGQVLYFDENGALSSTSYSTQBTNNL 223
QY 207 SQNATIFAANNRAYSTSAKNFEAVDNYLTADSWYRPKSILKDGKTTWTEGKDDFRPLLMA 266
DB 224 ---VTDFTKNNAAYSSTKASFELVDGYLTADSWYRPKLEAGTTWKASTEKDFRPLLMS 280
QY 267 WHPDFTETKRYNYNANKVVG-----IDKTYTAEQSADLTAAAEVLQARLEQKITSENKTK 322
DB 281 WHPDKDTQVAYLYNMTKALSNGEETKDVFTIENSQASLNAQAQILQKIEVIAANKSTD 340
QY 323 WLREAISAFVTKPQWNGSEKPYDDHLQNGALLFDNODTLTDPDTSNYRLNRTTPTNOT 382
DB 341 WLQSGIEAFVKDQKKNINSESPGKEHFKQGALLFVN-SDSTKWANSDRKLKNTATSY- 398
QY 383 GSLDSRFTYPNPDLGGYDFLLANDVDNSNPVQAEQLNMLHYLLNFGSIY--ANDADAN 440
DB 399 --IKNHKIVNGSD--GGYEFLLSNDIDNSNPVQAEMLNQLYYFNMWGOIVFGDKDKDAH 454
QY 441 FDSIRVADVNDVADLLQLISSDYLKAAYGIDKKNKANNHNVISIVAEWSNDTPYLLHDDGD 500
DB 455 FDGIRVADVNDVSDMLQLVSSYMKAAKYKNSSEARALANISILEAWSHNDPYVYVNEHT 514
QY 501 NLMNMNDKFRLSMLASLAKPL-DKESGLNPLIHNSLVD-----REVDDREVEVTPS 550
DB 515 AALSMDNGLRLSIVHGLTRPVTKGTGARNAMKDLINGGYFGLSNRAEVSYDQLGFAT 574
QY 551 YSFARAHDSQVDIIRDIKAEINPNSFGYSFTQBEIEQAFKIYNEDLAKTKDKYTHYV 610
DB 575 YLFEVRAHDSEVQTVIADIISKKIDPTDGTFTLQKQAFDIYNADMLKVDKEYTHSNI 634
QY 611 PLSYTLTLTNKSGIPRVYVGDMDPTDDGQYMANKTNYDAIESLLKARMKYVSGGQAMQNY 670
DB 635 PAAVALMLQTMGAATRVYVYDLYTDNGQYMAKSPYFDQITLLKARPKYVAGGQTSYIH 694
QY 671 QIG-----NGRILTSVRYGKALQKSD-KGDATTRTSGVGVVGMQNPFSL-DGKV 719
DB 695 NLAGDGVSSAKDNKEVLSVRYGQDLMSKTDTEGGKYGRNSGMLTIANNPLKLDAGET 754
QY 720 VALNMGAAHANOEYRALMVSTKDGVTATADADASKAGLVKRTDENGILYFLNDOLKGVA 779
DB 755 ITVNNGAHKKQAYEPPLLGTGKIGVSSINDSDTK---IVKTYDAQNLVFTADEIKGPK 811
QY 780 NPQVSGFLQVWVVGAAADQDQDIRVAASDTASTDGKSLH-QDAAMDSRVNFEFGSNFQSPA 838
DB 812 TVDMSGYLSVWVVPVGATDDQNVLAKPSTKAYKEGDKVYSSSALEAQVIYEGFSNFQDFV 871
QY 839 TKEEBEYTNVVIANNVDFKVSNGIITDFEMAPQVYVSTDGFQDLSVQNGVAFPTDRYDLGNS 898
DB 872 KEDSQYTNKLIANAADLFPKSWGITSFEIAPQVYSSKDGTFDLSIIENGVAFTDRYDFAMS 931
QY 899 KANKYGTADQLVKAIKALHAKGLKVMADWVPDQMTFFPKQEVVTVTRTDKFKPTAGSOI 958
DB 932 KNNKYGSKEDLRDALKALHKQGIQVIADWVPDQLYTLPGKEVVTATRTDTHGKVLDDTSL 991
QY 959 NHSLYVTDTKSGDDYQAKYGAFLDELKBYPELFTKKQISTGQAIDPSVKIKOWSAKY 1018
DB 992 VNKLVTNTKSSGNDFOAQYGAFLDKQLKLYPEIFKEVMEASGKTIDPSVKIKQWEAKY 1051
QY 1019 FNGSNILGRGADYVLSDOVSNKYFNVASDITFLPSSLLGKV-VESGIRYDGGKGIYNSSA 1077
DB 1052 FNGNTIQKSGSDYVLSLSD--GKLYFTVNDKGTFLPAALTGDTKAKTGFAYDGTGVYTYTT- 1108
QY 1078 TGDQVKASFIETAGNLYYFKDGYVMTGAOTINGANYFLENGTALRNTIYVDAQNSHY 1137
DB 1109 SGTQAKSOFVYNGKQYVFNKGILVTGEQITDGSNYFFLPNGVWFTDGRVKNKAGQSULV 1168
QY 1138 YANDGK-----RYENGYQFQGNDRWRYFKDGNMAVGLTVDGNVQYFDK 1180

```

Db 1169 YGSKGLTTQTCKWEKVVTKDSDGKEKFKYQF-----FKGGIMATGLTEVEGKEKIFYD 1222
QY 1181 DGVAQKDIIVTRDQKRVYFDQHNAGNAATNTFIADKTHGWHYLGKDGVAVTAQTVGKQK 1240
Db 1223 NGYQAQGVFPVTKDGLHLPFCGDSGERKYSGGF-EQDGNWYVANDKGYYVATGFTKVGQN 1281
QY 1241 LYFEANGQOVKGDFTVDSBGLYFYDVSDDMMWTDFTIEDKAGNWFYLGKGAAVTGAQT 1300
Db 1282 LYFNEKGQOVKNRF-----FQVGDA-----TYANNEGDLVRGAQT 1317
QY 1301 IRGQKLYFKANGQQVKGDIIVTKDQKIRYDADKSGEQVFNKTVKAADGKTVYVIGNDGVAV 1360
Db 1318 INGDELYFDESQKQVKGFPVNPDCGTSYDAITGVKL-----V 1356
QY 1361 DPS-VVKGQTFK-DASGALRFYNLKGQLVGTS---GWYETANHWDVYIOS-GKALTGBQT 1414
Db 1357 DTSLVVDGQTFNVDAKG-----VVTKAHTPGFYTTGDNWYFADSYGRNVTAQV 1406
QY 1415 INGOHLYPKEDGHQVKGQLVGCTDQKRVYDANSQDAFNKSVTVNGKTYFEGND-GTAQ 1473
Db 1407 INGOHLYFDANGROVKGGFVNTDGSRYFHWNTGDKLVSTFPFATGHRWYVYADRGNVV 1466
QY 1474 TAGNEPKQTFKDGSDIRFYSMEGQLVG-----SCWYENAOQ 1511
Db 1467 T-----GAQVINGQKL-FEDTDGQVKGAFATNANGSRYSYHWTGCKLVSTFFSGDNN 1520
QY 1512 WLYV-KNGKVLFGQVQSRVYPBENGIOAKGKAVRTSDGKIRYFDENSGMINTQW-K 1569
Db 1521 WYADAKGEVVVGEQTNGOHLFYDQTKQVKGATATPDGSGISYDVHTGKAINRWVK 1580
QY 1570 FYVQYVYFGNDG 1582
Db 1581 IPSGQWYVNAQG 1593

RESULT 8
T30857
Glucosyltransferase - Streptococcus salivarius
C:Species: Streptococcus salivarius
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30857
R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.
Infected. Immun. 63, 409-621, 1995
A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for pri
A:Reference number: Z20909; PMID:95122197; PMID:7822030
A:Accession: T30857
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1449 <SIM>
A:Cross-references: UNIPROT:Q55264; UNIPARC:UPI00000B166E; EMBL:L35495; NID:G662378; PID
A:Genetics:
A:Gene: gtfL

Query Match 37.5%; Score 3128.5; DB 2; Length 1449;
Best Local Similarity 42.9%; Pred. No. 2.5e-139; Mismatches 460; Indels 199; Gaps 31;
Matches 670; Conservative 231;

QY 1 MEKNYRFKHKYKKEKRWVTLVSASATM---LASALG-----ASVA---SADT 40
Db 1 MDKKVHYKHKYKQKQVVIATVGLSGLGAVSAVSLGTNDGVQDADEHTDATVAIPDITVD 60
QY 41 DTASDDSN-QAVVTGDDQTTNNQATDQTSIAAT----- 71
Db 61 GTVSNDDTTAAQDPPTTAVAAATNDVATDQATPTATDFTDITNTVANAADVATVATGTDRA 120
QY 72 -ATSEQASTDAATQAAAEQOTQGTASTDTAAOTTTNANEAKVWPTENENQGTDEMIL 130
Db 121 ATTNDTATNDTAVDTTN-NNTTDTTDTTDRAAATERRATGARGPTGRRATPVNGT 179
QY 131 AEAKNVATAESDIPSDLAKM-----SNVKQVDGKTYYYDQDGNVKNFAVS-VGDKIYFF 185
Db 180 NNANNTVTVNNDLPATNNVVDGFSHIKTNGKYVYEDDDGTIRKNVYLERIGGSQYFN 239

QY 186 DETGAYKDTSKVDADK-----SSSAVSQNAITFAANNRAYSTSAKNFEADVNDYLTADSW 239
Db 240 AETGELSNOKEYRDRKNGGTGSSADSTNTNTVNGDKNAFTGTTDKQLELDVGYETANTW 299
QY 240 YRPSILKDGKTWTESGKDDPRPLLMAMWPDTEKRYNTVYMNKV-VGIDKITYTAETSQA 298
Db 300 YRPEILLKDGKEWTASTENDKRPLLTVWMPSKAQASLYLNMKEOGLGNTQTYTSFSQT 359
QY 299 DLTAARALVQARIQKTSNTKWLREAIASFKVTQPOWNGESEK-PYDHLQNGALLF 357
Db 360 QMDQAALVEQKRIERIAREGNTDWRRTTKNFVKTPQGMNSTSENLDNDHLQGLL 419
QY 358 DNQTDLPDPTOSVRLNRTPTNQTGSLDSFTYNDPNDPLGQYDFLLANDVNSNPVQA 417
Db 420 NND5-RTSHANSDRYLLNRTPTQTGRHNPYTKDTSN--GGFEPELLANDIDNSNPVQA 476
QY 418 EQLNWLHYLLNFGSYIANDADANFDSIRVDADVNDVADLLQISSDYLKAAYGIDKKNKA 477
Db 477 EQLNWLHYIMNIGITGSGEDENFDGVRDAVDVNDVADLLQIASDYFKAYGADQSQA 536
QY 478 NNHVSIVEAWSNDPTPYLHDDGDMNMNDKFRJLSMLWSLAKPLDKRSLNPLIHNLSVD 537
Db 537 IKHL5ILEAWSHNDAYYNEDTKGAQLPMDPMHLALVYSLLRPIGNRSGVEPLISLND 596
QY 538 REVDDREVEVTPSYSPARAHDSVODIIRDIKAEINENSGYSFTQBEISQAQKYNED 597
Db 597 RSESGKSKRWNTAFVRAHDSVQSIIGQIKNEINPQSTGNTFTLDEMKAFAFYNKD 656
QY 598 LKTKDKKYTHVNPVLSYTLTLTNKGSIPRVYVGMFTDDGOVMANKTVNYDAIESLLKAR 657
Db 657 MRSANKQYQNIPSAYALMLTHKDTVPVYVGMFTDDGOVMAQKSPYDAIEFLKGR 716
QY 658 MKYVGGQAMQNYQINGE-----ILTSRVYKGGALKOSDKGDAATTRSGVGVNMG 709
Db 717 IRYAAGGDMKVYIGYNTNGWDAAGVLTSVRYGTGANSASDTGTAETRNQGMVIVSN 776
QY 710 QPNFSLDGKVALNMGAAHANOERYALMVSTKDGVAITYATDADAKAGLVKRTDENGVL 769
Db 777 QPALRLTSN-LTIINGGAHRNQYRPLLLTNDGVATYLLNDSAN--GIVKYTDGNGMLT 833
QY 770 FLNDLKGAVNPQVSGFLQVWPVGAADDQDIRVAASDTASTDGKSLHQDAMDSRVNFE 829
Db 834 F5ANEIRGIRNPQVDGYLWVVPVGASENQDVRVAPSKESGSLVYSNAAALDSQVIYE 893
QY 830 GFSNFPQSPATKEEYTNVIANNDKVFVSGWIGITDFEMAPQVYVSTDDGFLDSVTONGVAF 889
Db 894 GFSNFPQDPVQNPQVTKKIAENANLFKSWGITSFEFAPQVYVSSDDGSLDSVIONGVAF 953
QY 890 TDRYDLGSKANKYGTADQLVKAIKALHAKGLKVMADWVPDQMYTFPKQEVVTVTRDKF 949
Db 954 TDRYDIGMSKDNKYGSLADLKAALKSLHAVGISAIADWVPOIYNLPDGEVVVATRVN 1013
QY 950 GKPIAGSOINHSYVTDTKSSGDDYQAKYGGAFDELDELKEKYPPELTKKQISTGQADPSV 1009
Db 1014 GETDKGAILDHLAAKTRTFEGNDYQGYGAFDELDELKELYPQIFDRVQISTGKEMTDE 1073
QY 1010 KIKQWSAKYFNKSNILGRGADYVLSQVSNKYFNVASDTLFLPSSLLGKVVESGIRYDGK 1069
Db 1074 KITQWSAKYMGNTLDRGSEVYLKNGL-NGYGTNGGKVSLLPKVVGNSQNSGNGND 1132
QY 1070 G-----YIYNSSATGDQVKASFITE-AGNLYYFGKDGVMVTGAQTINGANYFF 1116
Db 1133 GSGKFEKRLFSVRVRYNN---GQYAKNAFTKNDGNVYVYFNSNGRMAVEKTIIDGKQFF 1189
QY 1117 LENGTALRNTIYTDAGQNSHYIYANDGKRYENGYQFGNDWRYFKDGNMAVGLTVDGNNVQ 1176
Db 1190 LANGVQLR-----DGYRQ-----NRRGQVF 1209
QY 1177 YFDKDGVOAKDKIIVTRDQKRVYFDQHNAGNA--TNTFIADKTHGWHYLGKDGVAVTGAQT 1235
Db 1210 YYDQNGV-----LNANGKQDPKPDNNNNSAGRNQFVQIGNVWYVYDNGKRVYTGQN 1262
QY 1236 VGKQKLYFEANGQVKGDFVTSDGKLYFYDVSDDMMWTDFTIEDKAGNWFYLGKGA 1295

Db 1263 INQELFFDNNGVQVKGRTV--NENGAIYYDANSGEWARRFAEIBPGVWAYFNNDGTAV 1321
QY 1296 TGAQTIRGQKLYFKANGQVKGDIKIRYDAKSGEQVFNKTVKAAADKTVVIGN 1355
Db 1322 KGSQNINGQDLFDQNGRQVKGALA--NVDGNLRYDVNSGELYRN----- 1365
QY 1356 DGVAVDPVSVKGTQFKDASGALRPFYNLKGLVGTSGWYETANHDDWYIY--SGKALTGEQT 1414
Db 1366 -----RFEIDG-----SWYFDGNGNAVKGWVN 1389
QY 1415 INQHLHYFKEDGHQVKGQVLTGDKGVYRYDANSQDAFNKSVTVN--GKTYIFGNDGTAQ 1473
Db 1390 INQNLFFDNNGKQIKGHLVR--VNGVVRYPDPNSGEMAVNRWVEVSPGWWVYFDGSGRQ 1448

RESULT 9
T30552
Glucosyltransferase N - Streptococcus salivarius (fragment)
C:Species: Streptococcus salivarius
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30552
R:Jaffe, R.I.
submitted to the EMBL Data Library, February 1998
A:Description: Streptococcus salivarius V1477 gcfn.
A:Reference number: Z20854
A:Accession: T30552
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1449 <JAF>
A:Cross-references: UNIPROT:O68542; UNIPARC:UPI00000810FD; EMBL:AF049609; NID:G2935545;
C:Genetics:
A:Gene: gcfn

Query Match 37.3%; Score 3114.5; DB 2; Length 1449;
Best Local Similarity 43.0%; Pred. No. 1.1e-138; Mismatches 464; Indels 197; Gaps 32;
Matches 671; Conserved 227; Mismatches 464; Indels 197; Gaps 32;

QY 1 MEKNVRFKMHKVKRWVLSVASATM---LASALG-----ASVA-----SADT 40
Db 1 MDKVKYKXHKVKKQVTVIAVGLSLGAVSAVSLGNDGVVQADEHTDVAIPDITVD 60
QY 41 DTASDDSN--QAVVTGDTNNQATDQTSIAA-----TATSQSASTDAATQASAA 90
Db 61 GTVSNDDTAAQDPTTAVATNDVATDQATPTATFDLTDTTINTVAANA VDTVATVGTDR 120
QY 91 EOTQGTASTDTAAQTNTN-----ANEAKWPTENENGFTDEMLA 131
Db 121 ATTNDTATNDTAVDTTNNNTTDTTNTNRAATTERRATGARRGPTGGRRATPVNGNTN 180
QY 132 EAKNVATAESDSIPSLAKM-----SNVKQVQDKYIYDODGNVKKNFVAVS--VGDKIYFPD 186
Db 181 NANNVTVVNNDLPATNVVTDGPHSHIKTINGKQYVEDDGTIRKNVYLERIGGSQYFNA 240
QY 187 ETGAYKTSKVDADK-----SSSAVSQNAATIFAANNRAYSTSAKNFEAVDNYLTADSWY 240
Db 241 ETGELSNQKEYRFDKNGGTGSSADSTNTNTVNGDKNAPYGTDDKDIELVDGYFTANTWY 300
QY 241 RPKSTLKDGKWTESGKODFRLLMAWHPDTEKKNYNNMKV--VGDKITVATTSQAD 299
Db 301 RPKELKDGKEWTASTDKRPLLVWVWPSKAIQASLYNMYKEQGLGNTQYTSFSSQTQ 360
QY 300 LTAALVQARTEQKITSENNTKLWREAIASFVKTPQWNGSEK--PYDDHLQNGALLPD 358
Db 361 MDQAALEVQKLEGRAREGNTDMLRTYIKNFVKTPQGNSTSENLDNNDHLQGGALLYN 420
QY 359 NOTDTPPTQSNRYLLNPTNTQTSLSRFTPNPNDPLGGYDFLLANDVNSNPFVQAE 418
Db 421 NDS--RTSHANSRYLLNPTTPTSTQCKHPKYTKDTSN--GGPEFLANDIONSFAVQAE 477
QY 419 QLNWLHYLLNFGSIYANDADANFDSIRVDADVNDVADLLQTSDDYLKAAYGIDKKNKAN 478
Db 478 QLNWLHYIMNIGTITGSGEDEFDGVVRDADVNDVNADLLQIASDYFKAKYGADQSDQAI 537

RESULT 10
T31098
probable dextranucrase (EC 2.4.1.5), extracellular - Leuconostoc mesenteroides

QY 479 NHVSVIVEAWSDDNTPYLHDDGDNLMNMNDFRLSLMLWSLAKPLDKESGLNPLIHNSLVDR 538
Db 538 KHLISILEASHNDAYNETDKGAQLPMDDPMLHALVYSLLRPIGNRSGVEPLISLNSLDR 597
QY 539 EYDDREVEVTPSVSPARAHDSVQDIIRDIKAEINPNSFGYSFTQEEIEQAQKYNEDL 598
Db 598 SESGKNSKEMANYAFVRAHDSVQSIIGQIKNEINPQSTGNTFTLDEMKAPEIYNKOM 657
QY 599 KXTDKKYTHYNPLSVTLALTNGKSIIPRVYVYGMFTDDGQYMANKTVNYDAISLAKARM 658
Db 658 RSANKQYTOYINPLPSAYALMLTHKDTVPVYVYGMFTDDGQYMAQKSPYDAIETLLKGR 717
QY 659 KTVSGQAQMYQIENGGE-----ILTSVRYGKGAALKQSKDKGADATTRTSGVGVWGNQ 710
Db 718 RYAAAGQDMKVNYIGVYNTNGWDAAGVLTSVRYGTGANSADTGTATRNQGNMAVIVSNQ 777
QY 711 PNFSLDGKVALNMGAAHANQYRALMVSTKQGVATYATDADAKAGLVKRTDENGVLVYF 770
Db 778 PALRLTSN--LTINMGAAHNRQYRPLLTNTDNGVATYLNDSAN--GIVKYTDCGNLTF 834
QY 771 LNDLKGVANPOVSGFLOVWVPVGAADODIRVAASDTASTDCKSLHQDAAMDSRVMEF 830
Db 835 SANEIRGINRNPQVDGYLAVWVPVGAENQDVRVAFSKEKNSSGLVYTESNAALDSQVIYEG 894
QY 831 FSNFQSFATKEEYTNVVIANNVDFVSMGIITDFEMAPQYVSVSTDCQFLDSVITQNGYAF 890
Db 895 FSNFQDFVPNPQYTNKKIAENANLFSKGIITSEFAPQYVSSDDSGFLDSVITQNGYAF 954
QY 891 DRYDLMGSKANKYGTADQLVKAIKALHAKGLKMAADWPDQMYTFPKQEVVTVTRTDKFG 950
Db 955 DRYDLMGSKANKYGSGLADLKAALKSLHAVGISAIDWPDQIYNLPDGEVWVATRVNNY 1014
QY 951 KPIAGSQNHSLYVTDTKSSGDDYQAKYGGALDELKELKYPFLTKQISTGQADPSVK 1010
Db 1015 ETKGAIIDHSLYAAKTRTFGNDYQGYGGAFLDELKRLYPOIFORVQISTGKRMVTDK 1074
QY 1011 IKQSAKTFNGSNIILRGADYVLSQVSNKYFNVASDTLFLPSSLLGKVVESGIRVDGKG 1070
Db 1075 ITKMSAKYNGTNIILDRGSEVYLKGL--NGYGTNGGKVSLLPKVGSNOSTNGDNQNGD 1133
QY 1071 -----YIYNSATGDQVKAFFITE--AGNLVYFGDKGYVMVTAQTINGANYFFL 1117
Db 1134 SGKPEKRLFSVRYRYNN--GOYAKNAFTKNDGNVYFDNSGRMAVGEKTIIDGKYFFL 1190
QY 1118 ENGTAIRNTIYDAQNSHYANDCKRYENGVOQFGNDWRYPKDGMAVGLTIVDGNVQY 1177
Db 1191 ANGVQLR-----DGYRQ-----NRRGQVYF 1210
QY 1178 FDKDGV--QAKDKIIIVTRDGKVRYPQHHNGNAATNTFIADKTHWYVYLGKDGVAVTGAQTV 1236
Db 1211 YDQNGVLSANGK----QDPKP----DNNNTSGRNOFVQIGNVWAYYDNGKRVIGHQNI 1263
QY 1237 GKQKLYFPANGQVKGDFVTSDEGLYFYVDSDGMWTDFTIEDKAGNMFYLGKDGAAVT 1296
Db 1264 NGQELFFDNNGVQVKGRTV--NENGAIYYDANSGEWARRFAEIBPGVWAYFNNDGTAVK 1322
QY 1297 GAQITRGQKLYFKANGQVKGDIKIRYDAKSGEQVFNKTVKAAADKTVVIGND 1356
Db 1323 GSONINGQDLFDQNGRQVKGALA--NVDGNLRYDVNSGELYRN----- 1365
QY 1357 GVAVDPVSVKGTQFKDASGALRPFYNLKGLVGTSGWYETANHDDWYIY--SGKALTGEQT 1415
Db 1366 -----RFEIDG-----SWYFDGNGNAVKGWVN 1390
QY 1416 NGQHLHYFKEDGHQVKGQVLTGDKGVYRYDANSQDAFNKSVTVN--GKTYIFGNDGTAQ 1473
Db 1391 NGQNLFFDNNGKQIKGHLVR--VNGVVRYPDPNSGEMAVNRWVEVSPGWWVYFDGSGRQ 1448

[illegible]

Search completed: February 11, 2006, 19:42:15
Job time : 44.8183 secs

Db 1712 PEIIVLNPNTF-HKKVNIINDSSSEYKWSSTEGSDFILVRYLEESNKKILOKIRI-KGIL 1769
QY 689 KOSDKGDA--RTSGVGVVMGNQPNFSLDGKVALNMGAAHNOEYRALMVSTKD 742
Db 1770 SNTQSFNKMISIDFKIKLISLGVIMSNFKSFENSELDRDLGFK-----IIDNK- 1819
QY 743 GYATVATDADASKA-CLVXRTDENGILYFLNDDLGKVANPOVSGFLQVWVPV-GAADDQD 800
Db 1820 ---TYYYDESKLVGLI--NINNSLFY-----DPIBFNLVTGWQTINGKKYYFD 1865
QY 801 IRVAASDTA--STDGKSLH--ODAAAMDSRVMEFSGPSNFOSFATKEEYTNVVIANNVDF 856
Db 1866 INTGAALTSYKLIINGKHFPYFNNDGVWMLGV-FKPGDGFYFAPANTQNNI----- 1915
QY 857 VSGITDFEMAPQYVSSSTDGQFLDSVIQNGYAFTRDYDLGMSKANKYGTADQLVKAIK-- 914
Db 1916 -----EGQAI--VYQSKFLTLNGKKYYFDNNSKAVTGWRIINNEKYY 1955
QY 915 -----ALHAKGLKVMADWPDQWYTFPKQEVTVTRDXFKGPIAGSQINHSLYVTDTKS 969
Db 1956 FNPNAIAAVGLQVIDN---NKYFNPDTAIIIS---KGWQTVNGSRY---YFDTDTAI 2004
QY 970 SGDDYQAKYGGAFDLDELKYPPELFTKKQISTGQAIDPSVKIKQWSA---KYE-----N 1020
Db 2005 APNGYKTIDGKHYPF-----SDCVVKIGVSTNGSNGFEYFAPANTY 2045
QY 1021 GSNILGRGADYVLSQVSNKYFNVASDTLFLPSSLGKVVESGIRYDGKGYIYN-----S 1076
Db 2046 NNIEGOAIVY-----QSKFLTLNGKKYYFDNN--SKAVTGWQIDSKYYFNTTABA 2097
QY 1077 ATGDQVKASFTIAGNLYFGKD-GWMTGAOTINGANYFFLENGTALRNTIYTDAGNS 1135
Db 2098 ATGWQ-----TIDGKKYYFNTTAAATGWQIDGKKYYFNTN-TAIASTGYTIINGKH 2150
QY 1136 HYIANDG-----KRYENGYO-----QFGND--- 1155
Db 2151 FYFNTDGMQIGVFGPNGFYFAPANTDANNIEGOAILYQNEFLTNGKKYYFGSDSKA 2210
QY 1156 ---WR-----YFKDGN--MAVGLTVDGNVQYFQDQVQAKKIIIVTRDGKRYFD-- 1201
Db 2211 VTGWRIINNKKYFNPNAIAAHLCTINNDKYFESYDGLQNGYITIERNN--FYFDAN 2268
QY 1202 -----QINGN-----AATNTFIADKTGHWYLLGKGVAVTG 1232
Db 2269 NESKMVTGVFKPNGFYFAPANTHNNNIEGOAIVYQNKFLT-LNGKKYYFDNDSKAVTG 2327
QY 1233 AQTGCKQKLYPEANGQVKGDFVTSDEGKLYFYDVSQDM---W-----TDTFI 1278
Db 2328 WQTDGKKYYFNLTAETAAATGWQITID-GKKYYFNLTAETAAATGWQIDGKKYYFNTTFFI 2386
QY 1279 EDKAGNWFYLGKGAAVTGAOTIRGKLYFKANGQOVKGDIVKGTGDKIRY-----DAK 1333
Db 2387 -----ASTGYTSINGKHFFYFNTDGMQIG-VFKGPNG-FEYFAPANTDAN 2429
QY 1334 SGEQ---VFNKTVKAADGKTYVIGNDGVAVDP-SVVKQGTFKDASGALRFYNLKGQLVG 1389
Db 2430 NIEGOAILYQNKFLTNGKKYYFGSDSKAVTGLRTIDGK-----KYYFNTNTAVAV 2480
QY 1390 SGWYETANHDWYIOSGKAL--TGEQTINGQHLFYKEDGHQVKGQLVGTGDKVRY--- 1444
Db 2481 TGM-QTINGKKYYFNTNTSIASGTGTIIISGKHFFYFNTDGMQIG-VFKGPDG-FEYFAPA 2537
QY 1445 --DANEGD-QAF---NKSVTVNGKTYFYFNGDGTQATAGNPKQGTFKDGSIDIRFYSMEGQL 1498
Db 2538 NTDANNIEGOAIRQNRFLYLHDNIYYFGNNSKAAT-----GWVTIDGN--RYF-FEPT 2589
QY 1499 VTSGWYENAOQWMLYVXNGKVLTLGLQTVGSRVYF-----DENGIOAKGKAVRTSD--- 1550
Db 2590 AMGANGYKTIIDKNKPYFRNGLPQIGVFKGSGNGFEYFAPANTDANNIE--GOAIRYQNRFL 2647
QY 1551 ---GKIRYFDENSGSMITNQWKFVGYQYFYFNGDGAIRYRG 1588
Db 2648 HLLGKIYYFGNNSKAV--TGMQTINGKYYFYFMPDPTAMAAAG 2686

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.
OM protein - protein search, using sw model
Run on: February 11, 2006, 18:57:35 ; Search time 214.546 Seconds
(without alignments)
5228.676 Million cell updates/sec
Title: US-10-797-821-37
Perfect score: 8349
Sequence: 1 MEKNVRFKMKVKRWTLS.....VYQYYFGNDGARIYRGNW 1590

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 2166443 seqs, 705528306 residues
Total number of hits satisfying chosen parameters: 2166443
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : UniProt_05.80.*
1: uniprot_eprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Length DB	ID Description
1	8349	100.0	1590	2 Q59983 9STRE
2	8256	98.9	1592	1 GTF2_STRDO
3	8221	98.5	1590	2 Q52663 9STRE
4	7713.5	92.4	1597	1 GTF1_STRDO
5	4663.5	55.9	1476	1 GTFB_STRMU
6	4392.5	52.6	1455	1 GTFD_STRMU
7	3636.5	43.6	1462	1 GTFD_STRMU
8	3604	43.2	1577	2 Q54178 STRGN
9	3527	42.2	1575	2 Q9LCH3_STROR
10	3502.5	42.0	1554	2 Q8KZL5_9STRE
11	3485.5	41.7	1454	2 Q69A94_LEUME
12	3464.5	41.5	1512	2 Q9WXJ5_9STRE
13	3410.5	40.8	1506	2 Q56CX8_9STRE
14	3325.5	39.8	1518	2 Q06000 STRSL
15	3282.5	39.3	2835	2 Q8G9Q2_LEUME
16	3245	38.9	1577	2 Q55265_STRSL
17	3214.5	38.5	1599	2 Q00599_STRSL
18	3128.5	37.5	1449	2 Q55264 STRSL
19	3114.5	37.3	1449	2 Q68542 STRSL
20	3088	37.0	1508	2 Q52224 LEUME
21	3087.5	37.0	1527	2 Q8KRE1_LEUME
22	3084.5	36.9	1522	2 Q6TXV4_LEUME
23	3083	36.9	1508	2 Q9EZHS_LEUME
24	3061.5	36.7	1527	2 Q9ZAR4_LEUME
25	3053	36.6	1477	2 Q9L466 LEUME
26	3031	36.3	1365	1 GTF5_STRDO
27	3025	36.2	1290	2 Q48756 LEUME
28	3003.5	36.0	1330	2 Q84CN4_LEUME
29	2994.5	35.9	1338	2 Q9WXJ4_9STRE
30	2974	35.6	1561	2 Q5SBM6_9LACO
31	2874	34.4	1463	2 Q5SBM6_LEUME

32	2808	33.6	1595	2 Q5SBM3 LACSK	Q5sbm3 lactobacill
33	2454	29.4	2057	2 Q9RE05 LEUME	Q9re05 leuconostoc
34	2411	28.9	1016	2 Q9LCJ7 LEUME	Q9lcj7 leuconostoc
35	2329.5	27.9	1781	2 Q4JCS4_LACRE	Q4jcs4 lactobacill
36	2312.5	27.7	1781	2 Q4JLC7_LACRE	Q4jlc7 lactobacill
37	2292.5	27.5	1772	2 Q5SBN3_LACRE	Q5sbn3 lactobacill
38	2284	27.4	1781	2 Q5SBL9_LACRE	Q5sbl9 lactobacill
39	2240	26.8	1772	2 Q5SBN0_LACRE	Q5sbn0 lactobacill
40	1759	21.1	591	2 Q8VUH3_STRMU	Q8vuh3 streptococc
41	1478	17.7	1619	2 Q5SBM0_LACRE	Q5sbm0 lactobacill
42	1460	17.5	1231	2 Q5SBN1_LACRE	Q5sbn1 lactobacill
43	959	11.5	522	2 Q6VV10_STRSA	Q6vv10 streptococc
44	825	9.9	374	2 Q6ZX19_9LACO	Q6zx19 lactobacill
45	678	8.1	2817	2 Q97K42_CLOAB	Q97k42 clostridium

ALIGNMENTS

RESULT 1
Q59983 9STRE PRELIMINARY; PRT: 1590 AA.
AC Q59983;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5).
GN Name=gtfI;
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1310;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OM2176;
RX MEDLINE=94146405; PubMed=8312602;
RA Sato S., Inoue M., Handa N., Aizawa Y., Isobe Y., Katayama T.;
RT "DNA sequence of the glucosyltransferase gene of serotype d
Streptococcus sobrinus.";
RL DNA Seq. 4:19-27(1993).
RN [2]
RP PROTEIN SEQUENCE.
RX MEDLINE=91224988; PubMed=1827439;
RA Mooser G., Hefta S.A., Paxton R.J., Shively J.E., Lee T.D.;
RT "Isolation and sequence of an active-site peptide containing a
catalytic aspartic acid from two Streptococcus sobrinus alpha-
glucosyltransferases.";
RL J. Biol. Chem. 266:8916-8922(1991).
DR EMBL; D13858; BAA02976.1; -; Genomic_DNA.
DR PIR; A39841; A39841.
DR HSP; P06653; 1HCX.
DR GO; GO:0047849; F:dextranucrase activity; IEA.
DR GO; GO:0046757; F:transferase activity, transferring glycosyl. .; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW binding_1; 2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Glucosyltransferase; Signal; Transferase.
FT SIGNAL 1 38
FT CHAIN 39 1590 glucosyltransferase-I.
SQ SEQUENCE 1590 AA; 175956 MW; C3C83A57CF3C2B0E CRC64;
Query Match 100.0%; Score 8349; DB 2; Length 1590;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1590; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MEKNVRFKMKVKRWTLSVASATMLASALGASVASADTDFTASDDSNQAVVTGQDTNN 60
DB 1 MEKNVRFKMKVKRWTLSVASATMLASALGASVASADTDFTASDDSNQAVVTGQDTNN 60
QY 61 QATDQTSIAATATSEQSASTDAADTQASAAEQTCGTATTASTDTAACTTTNANEAKVPTEN 120

DR	EMBL; D90213; BAA14241.1; -; Genomic_DNA.	
DR	HSSP; P06653; 1GVM.	
DR	InterPro; IPR002479; Cell wall bd put.	
DR	InterPro; IPR003318; Glyco_hydro_70.	
DR	Pfam; PF01473; CW_binding_1; 3.	
DR	Pfam; PF02324; Glyco_hydro_70; 1.	
KW	Dental caries; Glycosyltransferase; Repeat; Signal; Transferase.	
FT	SIGNAL 1 38 Potential.	
FT	CHAIN 39 1592 Glucosyltransferase-I.	
FT	REPEAT 1093 1142 1.	
FT	REPEAT 1158 1207 2.	
FT	REPEAT 1222 1272 3.	
FT	REPEAT 1287 1337 4.	
FT	REPEAT 1402 1451 5.	
FT	REPEAT 1514 1563 6.	
FT	REPEAT 1577 1592 7 (incomplete).	
FT	REGION 39 1044 Catalytic (approximate).	
FT	REGION 1093 1592 7 X tandem repeats.	
FT	REGION 1093 1592 Glucan-binding (approximate).	
SQ	SEQUENCE 1592 AA; 176168 MW; BC0A66D07931ECF CRC64;	
Query Match 98.9%; Score 8256; DB 1; Length 1592;		
Best Local Similarity 98.9%; Pred. No. 0;		
Matches 1574; Conservative 6; Mismatches 10; Indels 2; Gaps 1;		
QY	1 MEKNVRFKMHKVRWVTLVSASATMLASALGASVASADTTASDDSNQAVVTGDDTNN 60	
DB	1 MEKNVRFKMHKVRWVTLVSASATMLASALGASVASADTTASDDSNQAVVTGDDTNN 60	
QY	61 QATDQTSIAATATSEOSASTDAATDOASAAEQTGTASTDAAQTNNANEAKVPTEN 120	
DB	61 QATDQTSIAATATSEOSASTDAATDOASAAEQTGTASTDAAQTNNANEAKVPTEN 120	
QY	121 ENQGFDEMLAKNAVATAESDIPSLAKMNVKQVGGKYYYDODGNVKNFAVSVD 180	
DB	121 ENQGFDEMLAKNAVATAESDIPSLAKMNVKQVGGKYYYDODGNVKNFAVSVD 180	
QY	181 KIYFDETCAYKDTSKVDADKSSAVSQNATTFANNRAYSTSAKNFEADVNTADSWY 240	
DB	181 KIYFDETCAYKDTSKVDADKSSAVSQNATTFANNRAYSTSAKNFEADVNTADSWY 240	
QY	241 RPKSILKDGKWTESGKDFRPLLMAMPDTEKRNVMNMKVGVGIDKITYTAETSQADL 300	
DB	241 RPKSILKDGKWTESGKDFRPLLMAMPDTEKRNVMNMKVGVGIDKITYTAETSQADL 300	
QY	301 TAAAEVLQARIKQITSENNTKWLREAI SAFVKTPQWNGESEKPYDDHLQNGALLFDNQ 360	
DB	301 TAAAEVLQARIKQITSENNTKWLREAI SAFVKTPQWNGESEKPYDDHLQNGALLFDNQ 360	
QY	361 TDLTPTQSNRYLLNRTPTNQTGSLDSRFTYNNPNDPLGGYDFLLANDVDNSNPVQAEOL 420	
DB	361 TDLTPTQSNRYLLNRTPTNQTGSLDSRFTYNNPNDPLGGYDFLLANDVDNSNPVQAEOL 420	
QY	421 NMLHYLLNFGSIYANDADANFDSIRVDAVDNVDADLLQISSDYLKAAAYGIDKNNKNNH 480	
DB	421 NMLHYLLNFGSIYANDADANFDSIRVDAEDNDVADQLQISSDYLKAAAYGIDKNNKNNH 480	
QY	481 VSIVEAWSNDPTPYLHDDGDNLMNDKRLSMLSLAKPLDKRSLNPLIHNLSVDREV 540	
DB	481 VSIVEAWSNDPTPYLHDDGDNLMNDKRLSMLSLAKPLDKRSLNPLIHNLSVDREV 540	
QY	541 DREVEVTPSYSFARAHSEVDQIDIRDIKASINPNSFGYSTQEEIOAPKIYNEDLKK 600	
DB	541 DREVEVTPSYSFARAHSEVDQIDIRDIKASINPNSFGYSTQEEIOAPKIYNEDLKK 600	
QY	601 TDKKYTHNVPLSYLLLTNKGSIIPRVYVYGDMDFTDDGQYMANKTNNYDAIESLLKARMKY 660	
DB	601 SDKKYTHNVPLSYLLLTNKGSIIPRVYVYGDMDFTDDGQYMANKTNNYDAIESLLKARMKY 660	
QY	661 VSGQAMQNYQINGEILTSVRYGKALKQSDKGDATRTSGVVMGNQPNFSLDGKVV 720	
DB	661 VAGQAMQNYQINGEILTSVRYGKALKQSDKGDATRTSGVVMGNQPNFSLDGKVV 720	
QY	721 ALNMGAHAHQEYRALMWSTKDGVAITYATDADASKAGLVKRTDENGYLEFLNDDLKGVAN 780	
DB	721 ALNMGAHAHQEYRALMWSTKDGVAITYATDADASKAGLVKRTDENGYLEFLNDDLKGVAN 780	
QY	781 PQVSGFLQWVVPVGAADDDQIRVAASDTRASTDGKSLHQDAAMDSRVWFGSFFQSFATK 840	
DB	781 PQVSGFLQWVVPVGAADDDQIRVAASDTRASTDGKSLHQDAAMDSRVWFGSFFQSFATK 840	
QY	841 EBEYTNVVIANNVDKFSWGIITDFEMAPQVVSSTGQFLDSVIONGYATFTDRYDLGMSKA 900	
DB	841 EBEYTNVVIANNVDKFSWGIITDFEMAPQVVSSTGQFLDSVIONGYATFTDRYDLGMSKA 900	
QY	901 NKYGTADQLVKAIKALHAKGLKVMADWPDQMYTTPKQEVVTVTRTDKFGKPIAGSQINH 960	
DB	901 NKYGTADQLVKAIKALHAKGLKVMADWPDQMYTTPKQEVVTVTRTDKFGKPIAGSQINH 960	
QY	961 SLIYVTDTKSSGDDYQAKYGGAFDELKEKYPBLFTKKQMSGTQQAIDPSPVKIKQWSAKYFN 1020	
DB	961 SLIYVTDTKSSGDDYQAKYGGAFDELKEKYPBLFTKKQMSGTQQAIDPSPVKIKQWSAKYFN 1020	
QY	1021 GSNILGRGADYVLSQVSNKYFNVASDTLFLPSSLLGKVVESGIRYDGGKGYLYNSSATGD 1080	
DB	1021 GSNILGRGADYVLSQVSNKYFNVASDTLFLPSSLLGKVVESGIRYDGGKGYLYNSSATGD 1080	
QY	1081 QVKASFITEAGNLYYFGKDGVMVTGAQTINGANYFFLENGTALRNTIYTDAGNSHYAN 1140	
DB	1081 QVKASFITEAGNLYYFGKDGVMVTGAQTINGANYFFLENGTALRNTIYTDAGNSHYAN 1140	
QY	1141 DGKRY- -ENGVOQFGNDWRYFKDGNMAGVLTVDGNVQYFDKDGVOAKDKIIVTRDGKVR 1198	
DB	1141 DGKRYENENGYOQFGNDWRYFKDGNMAGVLTVDGNVQYFDKDGVOAKDKIIVTRDGKVR 1200	
QY	1199 YFDQNGHNAATNTEADTKGHMYYLGKDGVA VTGAQTVGKQKLYFEANGQVKGDFVTS 1258	
DB	1201 YFDQNGHNAATNTEADTKGHMYYLGKDGVA VTGAQTVGKQKLYFEANGQVKGDFVTS 1260	
QY	1259 EGKLYFYDSDGMDMTDTFIEDKAGNWFYLGKDGAAVTAQTIHQKLYFKANGQOVKGD 1318	
DB	1261 EGKLYFYDSDGMDMTDTFIEDKAGNWFYLGKDGAAVTAQTIHQKLYFKANGQOVKGD 1320	
QY	1319 IVKGTGDKIRYDAKSGQVFNKTVKAADGKTYVIGNDGVADVPVVKGTQTFKDSAGALR 1378	
DB	1321 IVKGTGDKIRYDAKSGQVFNKTVKAADGKTYVIGNDGVADVPVVKGTQTFKDSAGALR 1380	
QY	1379 FYNLAGQLVTGSGWYETANHDWVYIQSGKALTGEQTINGOHLVFKEDGHQVKGQLVGTGD 1438	
DB	1381 FYNLAGQLVTGSGWYETANHDWVYIQSGKALTGEQTINGOHLVFKEDGHQVKGQLVGTGD 1440	
QY	1439 GKRVYYDANSQDQAFNKSVTVNGKTYFFGNDGTAQTAGNPKQOTFKDGSDFRFSMEGOL 1498	
DB	1441 GKRVYYDANSQDQAFNKSVTVNGKTYFFGNDGTAQTAGNPKQOTFKDGSDFRFSMEGOL 1500	
QY	1499 VTGSGWYENAGQWLYVKNKGVLTLGLQTVGSRVYFDENGIOAKGKAVRTSGKIRYFDE 1558	
DB	1501 VTGSGWYENAGQWLYVKNKGVLTLGLQTVGSRVYFDENGIOAKGKAVRTSGKIRYFDE 1560	
QY	1559 NSGSMITNQWKEVYQYVYFGNDGARIYRGN 1590	
DB	1561 NSGSMITNQWKEVYRYYFGNDGARIYRGN 1592	
RESULT 3		
Q55263 9STRE		
ID	Q55263_9STRE PRELIMINARY; PRT; 1590 AA.	
AC	Q55263;	
DT	01-NOV-1996 (TRENBLrel. 01, Created)	
DT	01-JAN-1998 (TRENBLrel. 05, Last sequence update)	
DT	01-JUN-2003 (TRENBLrel. 24, Last annotation update)	
DE	GTF-1.	
GN	Name=Glucosyltransferase;	
OS	Streptococcus sobrinus.	
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;	
OC	Streptococcus.	

OX	NCBI_TaxID=1310;	
RN	[1]_TaxID=1310;	
RP	NUCLEOTIDE SEQUENCE.	
RC	STRAIN=ATCC 33478;	
RA	Sato S.;	
RT	"DNA and amino-acid sequences of water-insoluble-glycan synthetase	
RT	produced from Streptococcus sobrinus ATCC 33478.";	
RL	Kagoshima Daigaku Shigakubu Kiyo 16:23-29(1996).	
RN	[2]	
RP	PROTEIN SEQUENCE.	
RX	MEDLINE=91224988; PubMed=1827439;	
RA	Mooser G., Heita S.A., Paxton R.J., Shively J.E., Lee T.D.;	
RT	"Isolation and sequence of an active-site peptide containing a	
RT	catalytic aspartic acid from two Streptococcus sobrinus alpha-	
RT	glucosyltransferases.";	
RL	J. Biol. Chem. 266:8916-8922(1991).	
DR	EMBL; D63570; BAA09792.1; -; Genomic_DNA.	
DR	PIR; A39841; A39841.	
DR	HSSP; P06653; 1GVN.	
DR	GO; GO:009250; P-glycan biosynthesis; IEA.	
DR	InterPro; IPR002479; CW_binding.	
DR	InterPro; IPR003318; Glyco_hydro_70.	
DR	Pfam; PF01473; CW_binding_1; 3.-	
DR	Pfam; PF02324; Glyco_hydro_70; 1.	
SQ	SEQUENCE 1590 AA; 176058 MW; 9DF7A3F2C6B4FD43 CRC64;	
Query Match 98.5%; Score 8221; DB 2; Length 1590;		
Best Local Similarity 98.5%; Pred. No. 0;		
Matches 1566; Conservative 9; Mismatches 15; Indels 0; Gaps 0;		
QY	1 MEKVRFKVHKVKKRWVTLVSASATMLASALGASVASADTDPTASDDSNQAVVTGDTNN	60
DB	1 MEKVRFKVHKVKKRWVTLVSASATMLASALGASVASADTDPTASDDSNQAVVTGDTNN	60
QY	61 QATDQTSIAATATSEQSASTDAATDQASAAEQTGTASTDTAAQTNNANEAKVPTEN	120
DB	61 QATDQTSIAATATSEQSASTDAATDQASAAEQTGTASTDTAAQTNNANEAKVPTEN	120
QY	121 ENQGTDEMLAKAKVATAESDISPSDLAKSNVQVDGKYVYDQDGNVKNFAVSVD	180
DB	121 ENQGTDEMLAKAKVATAESDPSFSLAKSNVQVDGKYVYDQDGNVKNFAVSVD	180
QY	181 KIYVFDETGAYKDTSKVDADKSSAVSQNATIFAAANNRAYSTSAKNFEAVDNYLTADSWY	240
DB	181 KIYVFDETGAYKDTSKVDADKSSAVSQNATIFAAANNRAYSTSAKNFEAVDNYLTADSWY	240
QY	241 RPKSLKDGKTWTEGKDDFRPLLMAWPDPTETKRNYYNMVVGIDKTYTAETSQADL	300
DB	241 RPKSLKDGKTWTEGKDDFRPLLMAWPDPTETKRNYYNMVVGIDKTYTAETSQADL	300
QY	301 TAAELVQARIBOKITSENNTKWLREAIASFVKTPQWNGSEKPYDDHLQNGALLFDNQ	360
DB	301 TAAELVQARIBOKITTEQNTKWLREAIASFVKTPQWNGSEKPYDDHLQNGALKFDNQ	360
QY	361 TDLTPTDQTSNRLNRTPTNQTGSLDSRFTYNNPDPLGGLDPLANDVNSNPVVOAQL	420
DB	361 SLDLTPDQTSNRLNRTPTNQTGSLDSRFTYNNPDPLGGLDPLANDVNSNPVVOAQL	420
QY	421 NWLHYLLNFGSIYAKDADANFDSIRVDVNDVADLLQISSDYLKAAAGIDKNNKNANH	480
DB	421 NWLHYLLNFGSIYAKDADANFDSIRVDVNDVADLLQISSDYLKAAAGIDKNNKNANH	480
QY	481 VSIWEAWSDNTPYLLHDDGDLNMMNDKFRLSMLSLAKPLDKRSGLNPLIHNSLVREV	540
DB	481 VSIWEAWSDNTPYLLHDDGDLNMMNDKFRLSMLSLAKPLDKRSGLNPLIHNSLVREV	540
QY	541 DDREVTVPYSFARAHSEVDIIRDIKABINPNSFGYPTQEEIEQAFKIYNEDLKK	600
DB	541 DDREVTVPYSFARAHSEVDIIRDIKABINPNSFGYPTQEEIEQAFKIYNEDLKK	600
QY	601 TDKKYTHNVPLSYLLLTNKGSIIPRVYVYGDMDFTDGGYMANKTNYDAIESLLKARMKY	660
DB	601 TDKKYTHNVPLSYLLLTNKGSIIPRVYVYGDMDFTDGGYMANKTNYDAIESLLKARMKY	660

RESULT 4

CTF1_STRDO

ID	CTF1_STRDO	STANDARD;	PRT;	1597 AA.
AC	P11001;			
DT	01-JUL-1989	(Rel. 11, Created)		
DT	01-JUL-1989	(Rel. 11, Last sequence update)		
DT	13-SEP-2005	(Rel. 48, Last annotation update)		

QY	661 VSGGQAMQNYOINGEILTSVRYGKALKOSDKGDATTRTSGVGVVMGNPNFSLDGKVV	720
DB	661 VSGGQAMQNYOINGEILTSVRYGKALKOSDKGDATTRTSGVGVVMGNPNFSLDGKVV	720
QY	721 ALNMGAAHANOEYRALMVSTKDGVAITYATDADASKAGLVKRTDENGILYFLNDDLKGVAN	780
DB	721 ALNMGAAHANOEYRALMVSTKDGVAITYATDADASKAGLVKRTDENGILYFLNDDLKGVAN	780
QY	781 PQVSGFLQVWPVGAADDQDIRVAASDTASTDGKSLHQAADMSRVMEGFSNFQSFATK	840
DB	781 PQVSGFLQVWPVGAADDQDIRVAASDTASTDGKSLHQAADMSRVMEGFSNFQSFATK	840
QY	841 EEEVTNVVIANNVDFKYSWGTIDFEMAPQYVSSYSSDGGQFLDSVIQNGYAFTRDYDLGMSKA	900
DB	841 EEEVTNVVIANNVDFKYSWGTIDFEMAPQYVSSYSSDGGQFLDSVIQNGYAFTRDYDLGMSKA	900
QY	901 NKYGTADOLVKAIKALHAKGILKVMADWPDQMYTFPKQEVVTVTRTDKFGKPIAGSQINH	960
DB	901 NKYGTADOLVKAIKALHAKGILKVMADWPDQMYTFPKQEVVTVTRTDKFGKPIAGSQINH	960
QY	961 SLYVTDTKSGDDYQAKYGGAFDLDELKEXYDELFTKKQISTGQAIDPSVKIKQWSAKYFN	1020
DB	961 SLYVTDTKSGDDYQAKYGGAFDLDELKEXYDELFTKKQISTGQAIDPSVKIKQWSAKYFN	1020
QY	1021 GSNILGRGADVVLSDOVSNNKYFNVASDTLFLPSSILGKVVESGIRYDCKGYIYNSSATGD	1080
DB	1021 GSNILGRGADVVLSDOVSNNKYFNVASDTLFLPSSILGKVVESGIRYDCKGYIYNSSATGD	1080
QY	1081 QVKASFITEAGNLVYFGKDGVMVTGAOTINGANVFFLENGTALRNTIYTDAGQNSHYVAN	1140
DB	1081 QVKASFITEAGNLVYFGKDGVMVTGAOTINGANVFFLENGTALRNTIYTDAGQNSHYVAN	1140
QY	1141 DGKRYENGYQQFGNDWRYFKDGNMAVGLTTVDGNVYFDDKGVQAKDKIIIVTRDGKVRYP	1200
DB	1141 DGKRYENGYQQFGNDWRYFKDGNMAVGLTTVDGNVYFDDKGVQAKDKIIIVTRDGKVRYP	1200
QY	1201 DOHNGNAATNFTIADKTHWYLLGKGVAVTGAOTVGVKQKLYFEANGQVKGDFVTSDEG	1260
DB	1201 DOHNGNAATNFTIADKTHWYLLGKGVAVTGAOTVGVKQKLYFEANGQVKGDFVTSDEG	1260
QY	1261 KLYFYDSDGDMWTDFTIEDKAGNFWYLGKGGAAVTGAQTIRGQKLYFKANGQQVKGDIV	1320
DB	1261 KLYFYDSDGDMWTDFTIEDKAGNFWYLGKGGAAVTGAQTIRGQKLYFKANGQQVKGDIV	1320
QY	1321 KGTGDKIRYDAKSGEOVFNKTVKAADGKTYYIINNVAVDPSVVKGQTFKDAAGALRPFY	1380
DB	1321 KGTGDKIRYDAKSGEOVFNKTVKAADGKTYYIINNVAVDPSVVKGQTFKDAAGALRPFY	1380
QY	1381 NLKGOLVTGSGWYETANHDWYIOSGKALTGEOTINGOHLVFKEDGHQVKGOLVTRTDGK	1440
DB	1381 NLKGOLVTGSGWYETANHDWYIOSGKALTGEOTINGOHLVFKEDGHQVKGOLVTRTDGK	1440
QY	1441 VRYDANSGDQAQFNKSVTVNGKTYVYFGNDGTAQTAGNPKGQTFKDGSDIRFYSMEGQLVT	1500
DB	1441 VRYDANSGDQAQFNKSVTVNGKTYVYFGNDGTAQTAGNPKGQTFKDGSDIRFYSMEGQLVT	1500
QY	1501 GSGWYENAAQGWLVYKXGKVLTLGLQTVGSRVYFDENGIOAKGKAVRTSDGKIRYFDENS	1560
DB	1501 GSGWYENAAQGWLVYKXGKVLTLGLQTVGSRVYFDENGIOAKGKAVRTSDGKIRYFDENS	1560
QY	1561 GSMITNWKPFYVGYVYFGNDGABIRGWN 1590	
DB	1561 GSMITNWKPFYVGYVYFGNDGABIRGWN 1590	


```

Db 1261 FVTADGKLYFDVDSGDMWNTFTEDKAGNWFYLGKGNVTAQTIKGQKLYFKANGQ 1320
QY 1314 QVKGDIIVKGTGKIRYYDAKSGEYFNKTVKAAADGKTVYVIGNDGVAVDPSPVVGQTFKDA 1373
Db 1321 QVKGDIIVKADGKIRYYDAQTGEQYFNKSV-SVNGKTYVFGSDGTAQTQANPKGQTFKDG 1379
QY 1374 SCALPYNLKGQVLTGSGWYETANHDWYIQSGKALTGEOTINGQHLFEKEDGHQVKGL 1433
Db 1380 SGVLFYNLEGGYVSGSGWYETAEHFWYVKSGLVLTGAQTIGNQVYFKONGHQVKGL 1439
QY 1434 VTGTGKRYRYDANGDQAFNKSVTWNGKTYVFGNDGTAQTAGNPKGQTFKDGSDI-RFY 1492
Db 1440 VTGNDGKLYRYDANGDQAFNKSVTWNGKTYVFGSDGTAQTQANPKGQTFKDGSGVLRFY 1499
QY 1493 SMEGQVLTGSGWYENAOQGLVYKNGKVLTLGLQTVGSQVVPDENGIOAKGKAVRTSDGK 1552
Db 1500 NLEGGYVSGSGWYKNAQGLVYKDGKVLTLGLQTVGNQKVVYFDKNGIOAKGKAVRTSDGK 1559
QY 1553 IRYPDENGSMITNWKVFGYGGYVFGNDGARIYRGWN 1590
Db 1560 VRYPDENGSMITNWKVFGYGGYVFGSDGAAYVRGWN 1597

```

RESULT 5

```

GTFB_STRMU
ID GTFB_STRMU STANDARD; PRT; 1476 AA.
AC P09987; O69384; O69384; O69387; O69390; O69396;
DT 01-NOV-1988 (Rel. 09, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Glucosyltransferase-I precursor (BC 2.4.1.5) (GTF-I) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
GN Name=gtfB; OrderedLocusNames=SMU.1004;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=GS-5;
RX MEDLINE=87308013; PubMed=3040685;
RA Shiroza T., Ueda S., Kuramitsu H.K.;
RT "Sequence analysis of the gtfB gene from Streptococcus mutans.";
RL J. Bacteriol. 169:4263-4270(1987).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MT4239 / Serotype C, MT4245 / Serotype e, MT4251 / Serotype f,
RC MT4467 / Serotype e, and MT8148 / Serotype C;
RX MEDLINE=982331643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
RN [3]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=UAI59 / ATCC 700610 / Serotype c;
RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;
RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Prineaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,
RA Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental
RT pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
CC -!- FUNCTION: Production of extracellular glucans, that are thought to
CC play a key role in the development of the dental plaque because of
CC their ability to adhere to smooth surfaces and mediate the
CC aggregation of bacterial cells and food debris.
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-
CC fructose + (1,6-alpha-D-glucosyl) (n+1).
CC -!- SUBCELLULAR LOCATION: Secreted.

```

```

CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes both
CC water-soluble glucans (alpha 1,6-glucose). GTF-S1 synthesizes both
CC forms of glucans.
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
CC -!- SIMILARITY: Contains 10 cell wall binding repeats.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL; M17361; AAA88588.1; -; Genomic_DNA.
CC EMBL; D88651; BAA26101.1; -; Genomic_DNA.
CC EMBL; D88654; BAA26105.1; -; Genomic_DNA.
CC EMBL; D88657; BAA26109.1; -; Genomic_DNA.
CC EMBL; D88660; BAA26113.1; -; Genomic_DNA.
CC EMBL; D89577; BAA26119.1; -; Genomic_DNA.
CC EMBL; AE014940; AAN58705.1; -; Genomic_DNA.
CC PIR; B33135; B33135.
CC HSSP; P06653; 1H8G.
CC InterPro; IPR002479; Cell wall bd put.
CC InterPro; IPR003318; Glyco_hydro_70.
CC Pfam; PF01473; CW_binding_I; 4.
CC Pfam; PF02324; Glyco_hydro_70; 1.
CC Complete proteome; Dental caries; Glucosyltransferase; Repeat; Signal;
CC Transference.
CC SIGNAL 1 34 Potential.
CC CHAIN 35 1476 Glucosyltransferase-I.
CC REPEAT 1097 1130 A repeat.
CC REPEAT 1161 1210 1.
CC REPEAT 1225 1275 2.
CC REPEAT 1290 1340 3.
CC REPEAT 1355 1405 4.
CC REPEAT 1420 1470 5.
CC REGION 35 1051 Catalytic (approximate).
CC REGION 1097 1476 Glucan-binding (approximate).
CC REGION 1161 1470 5 X tandem repeats.
CC VARIANT 62 62 S -> T (in strain MT4239).
CC VARIANT 65 65 T -> I (in strain GS-5).
CC VARIANT 68 68 V -> A (in strain GS-5, strain MT4245,
CC strain MT4251, strain MT4467 and strain
CC MT8148).
CC VARIANT 78 78 Q -> P (in strain MT4251).
CC VARIANT 86 86 I -> S (in strain GS-5, strain MT4245,
CC strain MT4251, strain MT4467 and strain
CC MT8148).
CC VARIANT 89 89 S -> F (in strain MT4251).
CC VARIANT 168 168 K -> N (in strain MT4251).
CC VARIANT 276 276 S -> D (in strain GS-5, strain MT4467 and
CC strain MT8148).
CC VARIANT 399 399 N -> R (in strain MT4239).
CC VARIANT 474 474 I -> T (in strain MT4239).
CC VARIANT 512 512 K -> R (in strain MT8148).
CC VARIANT 519 519 F -> Y (in strain MT8148).
CC VARIANT 701 701 T -> I (in strain MT8148).
CC VARIANT 708 708 A -> V (in strain MT8148).
CC VARIANT 938 938 F -> L (in strain MT8148).
CC VARIANT 952 957 FGKPE -> YGTPVA (in strain GS-5, strain
CC MT4239 and strain MT4467).
CC VARIANT 963 964 SV -> NT (in strain GS-5, strain MT4239
CC and strain MT4467).
CC VARIANT 968 970 ADS -> VDG (in strain GS-5, strain MT4239
CC and strain MT4467).
CC VARIANT 1086 1086 A -> T (in strain MT4239).
CC VARIANT 1158 1158 S -> N (in strain MT4239).
CC VARIANT 1163 1163 H -> Y (in strain MT4251).
CC VARIANT 1168 1168 E -> K (in strain MT8148).
CC VARIANT 1182 1182 Y -> C (in strain MT8148).
CC VARIANT 1234 1234 A -> P (in strain MT4239).
CC VARIANT 1263 1263 R -> H (in strain GS-5 and strain
CC MT4467).

```

```
FT VARIANT 1263 1263 R -> P (in strain MT8148).
FT VARIANT 1264 1264 Y -> H (in strain GS-5, strain MT4239,
FT VARIANT 1272 1272 strain MT4467 and strain MT8148).
FT VARIANT 1272 1272 S -> G (in strain GS-5, strain MT4239,
FT VARIANT 1329 1329 strain MT4467 and strain MT8148).
FT VARIANT 1329 1329 H -> Y (in strain GS-5 and strain
FT VARIANT 1394 1394 MT4467).
FT VARIANT 1394 1394 Y -> H (in strain GS-5, strain MT4239,
FT VARIANT 1402 1402 strain MT4467 and strain MT8148).
FT VARIANT 1402 1402 S -> G (in strain GS-5, strain MT4239,
FT VARIANT 1459 1459 strain MT4467 and strain MT8148).
FT VARIANT 1459 1459 Y -> H (in strain MT4467).
FT CONFLICT 570 570 R -> A (in Ref. 1).
FT CONFLICT 800 817 AQQVRAASTAPSTDGK -> LIKMFALRLARPHQOMA
FT CONFLICT 1310 1310 (in Ref. 1).
FT CONFLICT 1476 1476 H -> L (in Ref. 1).
FT SEQUENCE 1476 AA, 165847 MW, 9C6E09F731B4CBFC CRC64;

Query Match 55.9%; Score 4663.5; DB 1; Length 1476;
Best Local Similarity 57.1%; Pred. No. 5.5e-207;
Matches 906; Conservative 206; Mismatches 339; Indels 135; Gaps 15;

QY 1 MEKVREKMKVKRWVTVLSVASATWLASALGASVASADTDTA-----SDSNQAVVTGQ 56
DB 1 MDKVKRYKRVKRWVTVSVASVMTLTLTSLGGLVKADSNESKSIQISNDSTSVVTANE 60
QY 57 TTNQATDTSIAATATSPQASASTDAATQASAAEQTQGTASTDTAAQTITNANEAKW 116
DB 61 ESN--VTTEVTSQBEAASQTNHTVTIISSTSVNPKVSNPYTVGETASNGEKL--- 115
QY 117 PTENENQGTDEMLAA-----KNVATAESDIPSD-----LAKMSNVQVQDKYYY 163
DB 116 ----QNOTTVTDKTSAAANNISQKTEADTDVIDDSNAANLQILEKLPNVKEIDKYYY 171
QY 164 YDQDGNVKNFAVSGDKIYYDETCAYKDTSKVDADKSSAVSQNATIPAAANNRAYSTS 223
DB 172 YDNGKVRTNFTLIADGKILHDETGAYTDTSDITVTK--DIVTTRSNLYKYNQVYDRS 229
QY 224 AKNFADVNDLTADSWYRPSILKDGKWTESGKDFRPLLMAMPDTEKRYNVYMNK 283
DB 230 AQSFVHVLTAESWYRPSYKILKDGKWTQSTQTEKDFRPLMTWPSQSTQRYVNYMA 289
QY 284 VVGIDKTYTAETSQADLTAAELVQARIEQKITSNNKTLREALSAFVKTPQWNGESE 343
DB 290 QLGINKTYDTSNQLQNLAAATIAQIEAKITLTKNTDMLAQITISAFVKTSAMNSDSE 349
QY 344 KPVDHLQNGALLFNQDITLTDPTQSNYLLNRTNPTQSLDSRFTYNDPDLGYPDL 403
DB 350 KPFDHLQNGAVLYDNEGLKTPYANSYRILNRTPTNQTGKKDPRT--ADNTIGGYEPL 407
QY 404 LANDVDSNPVQAELQNLWHLNFGSIYANDADANFDSIRVDADVNDADLLQISSDY 463
DB 408 LANDVDSNPVQAELQNLWHLNFGSIYANDADANFDSIRVDADVNDADLLQAGDY 467
QY 464 LKAAVYGDINKNKNANHNYSIVSAWSDNDTPYLHDDGNLMNNDKFRLSMLSLAKPLDK 523
DB 468 LKAAKGIIHKNDKAAANDHLSLEAWSDNDTPYLHDDGNLMNNDKFRLSMLSLAKPLNQ 527
QY 524 RSLGNPLNLSLVDRVDREVEVTPSYSFARAHSEVQDITRDIKAEINPNPSFGYPT 583
DB 528 RSGMNPNTNSLVNRTDNDNAETAAVPSYSFIRAHSEVQDLIRDIKAEINPNVVGYSPT 587
QY 584 OBEIEQAFKIYNEDLKTKTKYTHYNPLSYTLTLLTNKGSIPRVYVGMFTDDGOYMANK 643
DB 588 MEEKKAEIYNKDLLEKTKYTHNTALSVALLTNKSVPVRYVGMFTDDGOYMAHK 647
QY 644 TVNYDAIESLLKARMKYVSGGQAMQYQIENGCEILTSVRYKGLKQSDKGDATRTSGV 703
DB 648 TINYBAIETLLKARIKYVSGGQAMRNQVGNSEIITSVRYKGLKALCATDTGDRTRTSGV 707
QY 704 GYVMGNQNFSLDGK--VVALNKGAAHQAQRYALMVSTKDGCVATYATADAKAGLVKRT 762
DB 708 AVIEGNPNFLRKASDRVNVNNGAAHKQAVRPLLLTDDNGIKAYHSDQEA--AGLVRYT 765
```

RESULT 6

GTFC_STRMU

```
ID GTFC_STRMU STANDARD; PR: 1455 AA.
AC P13470; O69382; O69385; O69388; O69391; O69397; P05427;
DT 01-NOV-1988 (Rel. 02, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Glucosyltransferase-S1 precursor (EC 2.4.1.5) (GTF-S1)
DE (Dextranase) (Sucrose 6-glucosyltransferase).
GN Name=gTfC; OrderedLocusNames=SMU.1005;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
```

```
QY 763 DENGILYFLNDDLKGVANPOVSGFLOVWPVGAADDQDIRVAASDTASTDGKSLHODAA 822
DB 766 NDRGELIFTAADIKGYANPOVSGYLGWVPVGAADQDVRVAASTPSTDGKSVHQNAL 825
QY 823 DSRVMEPEGSFNSQSPATKBEETNNVIANVDFKVSAGITDFEMAPQYVSSSTGQFLDSV 882
DB 826 DSRVMEPEGSFNSQSPATKBEETNNVIANVDFKVSAGITDFEMAPQYVSSSTGQFLDSV 885
QY 883 IONGYAFTRDYDLGSKANKYGTADOLVKAIKALHAKGLKVMADWVPDQMYTPKQEVVT 942
DB 886 IONGYAFTRDYDLGSKANKYGTADOLVKAIKALHAKGLKVMADWVPDQMYTPKQEVVT 945
QY 943 VTRTDKFGKPIAGSQINHSLYVTDKSSGDDVQAKYGGAFDLDELKELPELTKQISTG 1002
DB 946 ATRVDKFGFVPGSKQSVLYVADSSKSKDQAKYGGAFLELOKAYPELFPARKQISTG 1005
QY 1003 QAIDPSVKIKQMSAKYFNGSNILRGADYVLSQVSNKYFNVA--SDTLFLPSSLGKVV 1060
DB 1006 VPMDPSVKIKQMSAKYFNGSNILRGADYVLSQVSNKYFNVA--SDTLFLPSSLGKVV 1065
QY 1061 ESGIRYDGGKYLNSSATGDOVKASFIETAGNLVYFKDGKGYMTGAQTINGANYFLENG 1120
DB 1066 QVGFSDYDGGKYYVYST--SGYQAKNTFISEGDKWYFPDNNGYMYTGAQSGVNYFSLNG 1124
QY 1121 TALRNTIYTDACNGSHYVANDGKRYENGYYQF--GNDWRVFKDGNMAVGLTTVDGNYQYFD 1179
DB 1125 LQIRDAILKVEDGTAYYVNDGRRYENGYYQFMSGVWRHFNNGEMSVGLTVIDGQVYFD 1184
QY 1180 KDGVOAKDKIIVTRDGKRYVDPQHNGNAATNTFIADKTGHYVYLGKDGVAVTGAQTVGKQ 1239
DB 1185 EMGYQAKGKFVTTADGKIRYFDKQSGNMYRNFRIENEEGKWLVLGEGDGAATVGSQTNGQ 1244
QY 1240 KLYPANGOOVKGDFVTSDEGKLYFDVDSGDMWTDTFIEDKAGWVYLGKDGAAVTGAQ 1299
DB 1245 HLYFRANGVQKGEFVTDYGRISYYSNDSGDIQIRNFRVNAQGGWYFPDNNGYAVTGAR 1304
QY 1300 TTRGQKLYFKANGQQVKGDIVKGTGKIRYIDAKSGEOVFNTKAAADGKTVYVIGNDGVA 1359
DB 1305 TINGQHLFRANGVQKGEFVTDYGRISYYSNDSGDIQIRNFRVNAQGGWYFPDNNGYAV 1364
QY 1360 VDPSSVVKQGTFKDASGALRFYNLKGLVTGSGWYETANHDWYVYI0SGKALTGEQTINGQH 1419
DB 1365 V-----TGARTINGQH 1375
QY 1420 LYFKEDGHQVKGQLVTGTDGKRYVYDANGSDQAFNKSVTVNGKTYFNGNDGTAQTAGNPK 1479
DB 1376 LYFRANGVQKGEFVTDYGRISYYSNDSGDIQIRNFRV----- 1413
QY 1480 GQTFKDGSDIRFYSMEGQLVTGSGWYENAGQWLVV--KNGKVLTLGLQTVGSORVYFDENG 1538
DB 1414 -----RNAQGGWYFPDNNGYAVTGTARTINGQHLRYPRANG 1447
QY 1539 IQAKGKAVRTSDGKIRYFDENSGSMI 1564
DB 1448 VQVKGFEVTDYGRISYYSNDSGERSV 1473
```

RESULT 6

GTFC_STRMU

```
ID GTFC_STRMU STANDARD; PR: 1455 AA.
AC P13470; O69382; O69385; O69388; O69391; O69397; P05427;
DT 01-NOV-1988 (Rel. 02, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Glucosyltransferase-S1 precursor (EC 2.4.1.5) (GTF-S1)
DE (Dextranase) (Sucrose 6-glucosyltransferase).
GN Name=gTfC; OrderedLocusNames=SMU.1005;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
```


This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.

```

-----
DR EMBL; M29296; ABA26895.1; -; Genomic DNA.
DR EMBL; D88653; BAA26103.1; -; Genomic DNA.
DR EMBL; D88656; BAA26107.1; -; Genomic DNA.
DR EMBL; D88659; BAA26111.1; -; Genomic DNA.
DR EMBL; D88662; BAA26115.1; -; Genomic DNA.
DR EMBL; D89979; BAA26121.1; -; Genomic DNA.
DR EMBL; AE014932; AAN58619.1; -; Genomic DNA.
DR HSP; P06653; IGVN.
DR InterPro; IPR002479; Cell_wall_bd_pnt.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF02324; CW_binding_1; 4.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Complete proteome; Dental caries; Glycosyltransferase; Repeat; Signal;
Transferase.
SIGNAL 1 ? Potential.
CHAIN ? 1462 Glucosyltransferase-S.
REPEAT 1232 1295 1.
REPEAT 1296 1359 2.
REPEAT 1360 1423 3.
REGION 1232 1423 3 X 63 AA approximate tandem repeats.
FT y -> H (in strain GS-5, strain MT4239,
FT strain MT4245, strain MT4251, strain
FT MT4467 and strain MT8148).
FT I -> V (in strain GS-5, strain MT4239,
FT strain MT4245, strain MT4251, strain
FT MT4467 and strain MT8148).
FT K -> E (in strain MT4467).
FT A -> S (in strain MT4239 and strain
FT MT4245).
FT A -> T (in strain MT4251 and strain
FT MT8148).
FT T -> I (in strain MT4239 and strain
FT MT4245).
FT A -> V (in strain MT4239, strain MT4245
FT and strain MT8148).
FT A -> S (in strain GS-5 and strain
FT MT4467).
FT A -> V (in strain MT4245).
FT A -> T (in strain GS-5, strain MT4239,
FT strain MT4245, strain MT4251, strain
FT MT4467 and strain MT8148).
FT V -> L (in strain MT4239).
FT D -> N (in strain MT8148).
FT E -> D (in strain MT4239, strain MT4245
FT and strain MT4251).
FT D -> N (in strain MT4239, strain MT4245
FT and strain MT4251).
FT Q -> H (in strain MT4245).
FT D -> N (in strain MT4239 and strain
FT MT4251).
FT E -> K (in strain MT4239).
FT V -> F (in strain MT4239).
FT F -> L (in strain MT4239, strain MT4251
FT and strain MT4467).
FT KKKTQ -> EKEYTL (in strain MT4251).
FT A -> S (in strain MT4239).
FT TDQGS -> ADKGNDS (in strain MT4251).
FT TDQGS -> ADKGN (in strain MT4239 and
FT strain MT4245).
FT T -> A (in strain GS-5, strain MT4239,
FT strain MT4245, strain MT4251, strain
FT MT4467 and strain MT8148).
FT D -> Y (in strain MT4251).
FT E -> K (in strain MT4245 and strain
FT MT4251).
FT LG -> IR (in strain MT4251).
FT G -> R (in strain MT4245).

```

```

FT VARIANT 1080 1080 G -> R (in strain MT4239).
FT VARIANT 1142 1142 H -> Q (in strain GS-5).
FT VARIANT 1198 1198 S -> N (in strain MT4239).
FT VARIANT 1220 1220 Y -> C (in strain MT4251 and strain
MT4467).
FT VARIANT 1280 1280 F -> L (in strain MT4467).
FT VARIANT 1282 1282 Q -> P (in strain MT4245).
FT VARIANT 1290 1290 K -> T (in strain MT4245).
FT VARIANT 1311 1311 N -> D (in strain MT4245).
FT VARIANT 1403 1403 G -> D (in strain GS-5 and strain
MT4467).
FT VARIANT 1425 1425 G -> R (in strain GS-5).
FT VARIANT 1449 1449 R -> K (in strain MT4467).
FT CONFLICT 1428 1462 RYDKNSGMVYKVKVHTLNGRRIGIDRWGIARY -> VY
R (in Ref. 1).
SQ SEQUENCE 1462 AA; 163388 MW; CE4A279C4D708645 CRC64;

Query Match 43.6%; Score 3636.5; DB 1; Length 1462;
Best Local Similarity 48.2%; Pred. No. 1.7e-159;
Matches 732; Conservative 247; Mismatches 412; Indels 129; Gaps 27;

QY 1 MEKNVRFQKHVKKRWVLSVASA--TMLASALGASVAS-----ADTDRAS 44
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 METKRYRYKVKKHVVTVASGLITLTGTTLGSSVSAETEQQTSKDVTKSDDKAA 60

QY 45 DDSNQA---VVTGDTNNQATDQTSIAATATS-EQSASTDAATDOASAAEQTGTFAST 100
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 SESSQTDAPKTKQAQTEQQAQSNVADTSTSTKETPSONITTTQANSDDKVTNTKSE 120

QY 101 D--TAAQTNTNANAKVPTENENOGFTDEMLAEAKVAT-----ASSDSIPSLAKMS 152
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 EAQTSERTKQAEQAATA-----SSQALTQAKAELTKQRTAAQENKPNVDAALP 172

QY 153 NVKQVDGKYYYYDQGNVKXNPAVSGDKIYYFDE--TGAYKDTSKVDADKSSSAVSQAT 211
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 173 NVKQIDGKYYVIGSDGQPKKFNALTANNKLVYFDKNTGALTDTDSYQFKQGLTKLND-- 230

QY 212 IFAANNRAYSTSAKNFEADVNYLTADSWYRPSILKDGKTWTESGKDFRPLLMWWPDT 271
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 231 -YTHNQVNFNTSLETIDNYVTADSWYRPFKDLKNGKWTASSESLRPLLMWSWPDFK 289

QY 272 ETKRNYVMNMKV-VGIDKTYTAETSOADLTAAAEVQARI EOKITSENNTKWLREISA 330
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 290 QTQATAYLNMNQOGLGTGENTYADSSQESLNLAQTVQVKIETKISQQTQWLRLDINS 349

QY 331 FVKTPQPNWNSQTESDTSAGEKDLQGGALLYSN--SDKTAYANSDYRLNRTPTSQTG--- 405
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 350 FVKTPQPNWNSQTESDTSAGEKDLQGGALLYSN--SDKTAYANSDYRLNRTPTSQTG--- 405

QY 387 SRFTYNNPDLGGYDFLLANDVNSNPVQAEOLNWLHLLNFGSIYANDADANFDSIRV 446
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 406 -KPKYFEDNSGGYDFLLANDVNSNPVQAEOLNWLHLLNFGSIYANDPEANFDGVRV 464

QY 447 DAVDNVDADLLQISDYLKAAVGDKNKNANNHVSIVEAWSNDNTPYLHDDGDMLMND 506
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 465 DAVDNVDADLLQIASDYLKAAVGDVSKSEKNAHLSILEAWSNDNDPQYKDTKGQALPD 524

QY 507 NKFRSLMWSLAKPLDK-----RSGLNPLIHNSLVDREVDDREVETVPSYSFARAH 558
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 525 NKLRSLLYALTRPLEKDNKNEIRSGLEPVITNSLNNRGAEGKNSEMANIYIFRAHD 584

QY 559 SEVDQIDIRDIKAEINPNSFGYSFTQEBIQAFKYNEDLKKTKKTYHYNVPYSYTL 618
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 585 SEVDQVIAKIIKAQINPKTDGLTFTLDELKQAFKYNEDMRQAKKYYTQSNIPYAYLML 644

QY 619 TNKSGIPRVYGDMPDQGVYKNTVNYDAIESLLKARMKYVSGGQAMQ-NYQICNGS- 676
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 645 SNKDSITFLYGDWYSDGQWATKSPYDAIDTLKARIKYAAGQDMKITVYVSGDKSH 704

QY 677 -----ILTSVRVYKGAALKQSKDGDATRTSGVGVVMGNQPNFSLD-GKVVALNMGAAH 728
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 705 MDWDVTGLTVSVRYGTGANEATDQGSATKQGMVITSNPNPSLKNQNDKVINMGTAH 764

```

```
QY 729 ANOEYRALMSTVSKDGVATYATADASKAGLVKRTDENGVLFLNDDLKGVANPOVSGFLQ 788
Db 765 KNOEYRPLLLTDDGLTSTVSDAAK--SLYRTDKDGBELVFDASDIQIYLNLPQVSGYLA 822
QY 789 VVVPVGAADQDIRVAASDASTADGKSLHQAAMDSRVMFEGSNFQSFATKEEYTVNV 848
Db 823 VVVPVGAASDQVRVAASKANATGVYESSALDSQLIYEGFSNFQDFVTKDSYTNKK 882
QY 849 IANNVDKFKVSGITDEMAPOVSVSDGQFLSVIONGVAFTRDYDLGMSKANKYGTADQ 908
Db 883 IAQNVQLFKVSGVTSFEMAPQVSVSDGSLFSDIIQNGVAFEDRYDLAMSKNKYGSQD 942
QY 909 LVKATKALHAKGLKGMADWVPQMTTFPKQEVVTVTRTDKFGKPTAGSINHSLYTDFK 968
Db 943 MINAKVALHSGIQVADWVPQIYNLPKGVVTVATRVNDYGEYRDKSEIKNTLYAANT 1002
QY 969 SSGDDYQAKYGGAFDELKEKYPELFTKKQISTGQAIDPSVKIKOWSAKYFNNGSILGR 1028
Db 1003 SNGKDYQAKYGGAFSELAAPYIFNRTOISNGKKIDSEKITAOKAYFNGTILGR 1062
QY 1029 ADYVLSDOVSNKYFNVASDTLFLPSLLGKVVEGIRYDGKGYIYNSSATGQVKAFTT 1088
Db 1063 VGYVLKDNASDKYFELKGNQTLPLKQMTNKEASTGFVNDGNGMTFYST-SGYOAKNSFVQ 1121
QY 1089 EA-GNLYYFGKDGVMVTGAQTINGANYFELENGTALRNTIYTDAGNSHYANDGKRYEN 1147
Db 1122 DAGKNWYFDDNHNHVMYQHLNGEVOYFELNSGVQLRESFLENADGSKNYFGLGNRYSN 1181
QY 1148 GYQOFEND--WRYF-KDGNMAGLTVTCNVGYFDKDGVOAKDKIIVTRDGVRYFDOHN 1204
Db 1182 GYSPDNDKWRVYFASGVMAVGLKTINGTQYFDQDGVQVKGANITGSDGKKRFDGGS 1241
QY 1205 GNAANTFTIADTGHYLYGKDGVAVTGAQTVGKQKLYPEANGQVKGDFVTSDEGKLYF 1264
Db 1242 GNMAVNRFPANDKGDWYVLSNDSGIALGVQTINGKTYTFGQDGKQIKGIIT-DNGKLY 1300
QY 1265 YVDGSDMWTFIEDKAGNFYLGKDAVNTGAQTIRGQKLYFKANGQVKGDIYKGTD 1324
Db 1301 FLANSGLARNIIFATDSQNNWYFSGDVGAVTGSQTIAKKLYFASDGKQVKSFTV-YN 1359
QY 1325 GKIRYYDAKSBOVENKTVKAADGKTYVIGNDGVAVDPSVVKGTQFKDASGALRFYNLKG 1384
Db 1360 GKHYHADSGLQVNRFEADKG----- 1383
QY 1385 QLVTSQGWYETANHDWVYIQS-GKALTGEQTINGQHLFYKEDGHQVKGQLVGTGDKVRY 1443
Db 1384 -----NWYLLDSNGEALTSQORINGRVFFTRREGKQVKGDVAVDERGLLY 1429
QY 1444 YDANSGDOAFNKSVTIV-NGK 1462
Db 1430 YDRNSGNMVKVVTLANG 1449
```

RESULT 8

```
Q54178_STRGN PRELIMINARY; PRT; 1577 AA.
ID Q54178_STRGN PRELIMINARY; PRT; 1577 AA.
AC Q54178; Q54247;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Glucosyltransferase.
GN Glucosyltransferase.
OS Streptococcus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1302;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Challis;
RX MEDLINE=96157084; PubMed=8586195;
RA Vickerman M.M., Sulavik M.C., Clewell D.B.;
RT "Molecular analysis of Streptococcus gordonii glucosyltransferase
    phase variants.";
```

```
RL Dev. Biol. Stand. 85:309-314(1995).
RP [2]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=Challis;
RA Minick P., Vickerman M.;
RL Submitted, (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; U12643; AAC343483.1; -; Genomic_DNA.
DR PIR; B41898; B41898.
DR HSSP; P06653; 1HCX.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 5.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1577 AA; 177805 MW; 5AB0328DC5E08D18 CRC64;

Query Match 43.2%; Score 3604; DB 2; Length 1577;
Best Local Similarity 44.9%; Pred. No. 6e-158;
Matches 756; Conservative 252; Mismatches 457; Indels 220; Gaps 35;

QY 1 MEKNVREKMHKVKKEWVTLVSASATMLAS--ALG-----ASVASA 38
Db 2 MEKKVHYKMKHVKKQWVAIAVTSLALLVAPKALGLESGLVIYADDANQVTVNVEQSASQVK 61
QY 39 D-----TDTASDDSN-----QAVVTGDTQTTNNQATDQTSIAATA----- 72
Db 62 DSEQTTSKATDSSQLEKVEQASSSKETVQASATNPANEQTQODKEVETSRDTSRHE 121
QY 73 -----TSQSASTD--AATDQASAEQTGT--TASTDTAA-----OTTNANEA-- 113
Db 122 LTQKTSDDSSKSGSQEPKQVADQAESTDKTQALQAKQDSRANDQEETENVAKATVD 181
QY 114 KWPFTENE-----NQGFDEML---AEAKNVATAESDSIPSDLAKMSNVKQVDGKY 161
Db 182 KLIATPKKERLPEPAQRKESITEKMLAAQAAPVNTHEHDDV-----LAHIKTIIDGKK 235
QY 162 YYYDQGNVKNKFAVSGDKIYFDF-ETGAYKDTSKVDADKSSASVQNAVTFPAANNRAY 220
Db 236 YVYDQDGTVKNFVAVELNGKILYFDAETGALVDSNEYQFOQGTSSINNE---FTQKNARY 292
QY 221 STSAKNFAVNYLTADSWYRPSKILKDGKTWESGKDDFPRLLMAMPDPTETKRYNY 280
Db 293 GTTDKDIETVDGYLTADSWYRPKFILKDGKTWASTETDLRPLLMAMPDKRTQINLYN 352
QY 281 MNKV-VGIDKTYTAETSQADLTAAELVQARIEQKITSENNTKWLREISAIVKTPQOWN 339
Db 353 MNQENLGIG-APESKTEQVLLTNVAVQVQVQRIEERISKEGDKTLRLTMSAEVKTQPNW 411
QY 340 ---GESEKPYDDHLQNGALLFDNOTDLTPDTQSNYRLNLTPTNTQTSGLDSRFTYNDP 395
Db 412 IKTSETTGTNKHLOGGALLYTN-SDKTSANSRYRLNLTPTTSQTGT-PKYFTDKSN- 468
QY 396 PLGGYDFLLANDVDNSNPVQAEQLNLWLLHYLLNFGSIYANDADANFDSIRVDAVDNVDAD 455
Db 469 --GGYEFLLANDFDNSNPAVQAEQLNLWLLHYMNFGSIIVANDPTANFDGVVDAVDNVDAD 526
QY 456 LLQISSDYLKAAYGIDKKNKNNHVSIVBEAWSNDPTLHDDGNLMMNMKNKFLSLMLW 515
Db 527 LLQIASDYFKSRYKVGSEEEALHLSILEAWSNDPDYNDKDTKGAOLAIDNKLKLSLY 586
QY 516 SLAKPLDKRSGNLPLIHNSLVDRVDDREVETVPSYFARAHDSVQDIIRDIKAENP 575
Db 587 SFMRKLSIRSGVEPTIYNSLNDSTRSTENKNGERTANYIFRAHDSVQTVIADIIRENINP 646
QY 576 NSFGYSFTQEEIEQAFTIYNEDLKKTKYTHYNVPLSTLLLTNKGSIIPRYVYDGMFTD 635
Db 647 NTDGLTFTMDELKQAFKIYNEDEMRKADKKYQFNIPTAHALMWSNKDSITRYYGDLVTD 706
QY 636 DGQWANKTVNYDAIESLLKARKMYVSGGQAMQYQIG-----NGEILTSVRVYVK 685
Db 707 DGQWMEKKSYPYHDAIDALLRARIKYVAGQDMKVTYMGVPPREADKWSYNG-ILTSVRVYGT 765
```


Db 588 TDGFTFTLDQLKQAFDINYNKDMNSVEKHYYTHYNI PAAYAVMLSNMESVTRVYVYGLFTDD 647
QY 637 GQYMAKNTVNDYDAIESLLKARKMYVSGQAMQNYQ-----IGNGEILT 679
Db 648 GQYMETKSPYYDAINTLLRARIRYAAGGTMEHKAFTYPSAAMKAKNPDSGSLVGNSEVLV 707
QY 680 SVRYGKALKQSD-KGDATTRTSVGVVGNQPNFSLD-GKVVALNMGAHAHQEYRALM 737
Db 708 SVRFQGVMSADMTGQGLAKTSGMPSLIANNPELELDANEIEIKVNVGKIHAQAYRPLL 767
QY 738 VSTKDGATYATDADAKAGLVKRTDENGLYFLDNDLKGAVNPQVSGFLQVVPVGAAD 797
Db 768 LTTDKGLQKLYNDSNTN---LTKVADKCGFITFKGEIKGYKQVEVNGVLSVVPVGAKS 824
QY 798 DQDIRVAASDTASTDG-KSLHQDAWDSRVMEFEGSNFOSFATKBEETVNVVIANVVKF 856
Db 825 DQDIRVAASSTKANAKGDKSYTASQALDSQLIIEGFSNFQDFQKDAQYTNKKIAENTDLF 884
QY 857 VSWGILTFEMAPQVYSTDQGLDSVIONGYAFTDRYDLGMSKANKYGTADOLVKAIKAL 916
Db 885 KAWGVTSEFMAPOYVSATDGTFLDSIIQNGYAFSDRYDLAMS KNNKYGSKEDLANALKAL 944
QY 917 HAKGLKWMADWYPDQWYTPPKQEVTVTRTDFGKPIAGSQINHSILYVTDKSSGDYQA 976
Db 945 HAAGTQAIADWPDQIYQIPGKEVTVASRVNRYGRVKIDQPNVKNKLYLANTHSSGKDFQA 1004
QY 977 KYGGAFLDELKBYPELFTKKQISTGQAIIDPSVKIKQMSAKYFNGSNILGRGADYVLSQ 1036
Db 1005 KYGGEFLAELQYPEMFATKMIISTGKPIDSSVLKKEWSAQYFNGTNVLGRGTDVYLSDE 1064
QY 1037 VSNKYPNVAVDLFLPSSLLG-KVVEGIRYDGKGYIYNSSATGQOVKASFITEAGNLYY 1095
Db 1065 GTGKYFTVNERGEFLPAVLGTGDEAKTGFYNDGKGMTYFTTA-GSQAQSDFTVAGNTY 1123
QY 1096 FCKDGYMTVGAOTINGANFFLENGTALRNTTYTDAQGNSHYVANDGKRYENGYQOFGND 1155
Db 1124 FYDTGHWITGNGINTKFFYFLPNGVMLKDAVWEDDRGSRVYVYKGTVMYKGSR---NNE 1180
QY 1156 W-----RYFKD-GNMAVGLTTVDGNVQYFDKDGVOAKDKIIIVTRDGKRYFDQ 1202
Db 1181 WFAWTDKQGLRFRHFDNFYGFMSGLVTHGNVQYVDEBGFQVKGDFVTDKAGQTRYFDK 1240
QY 1203 HNGNAATNTFIADKTHWYLGKDGVAVTGAOTVKGOKLYPEA-NGQOVKGFVTSDEBK 1261
Db 1241 NTGNLVKGOFF-NQNGHWYYSDDQGLIAKGAQTIRGOKLYFPAKTGAQVKGDFVTDKGN 1299
QY 1262 LVFYDVDSGDMWTDFTIEDKAGNWFYLGKDGAAVTAOTIRGOKLYFKA-NGQOVKGDIV 1320
Db 1300 TFFYSDDTGLAVSTFFSTGNNANFYADENGHVAKGEKTINGOKLYFDTKTGQQAQGRPV 1359
QY 1321 KGTDGKIRYDAKSQEVENK--TVKAADGKTYVIGNDGVAVDPSVYVKGQTFKDGASGALR 1378
Db 1360 RDAKG-LRFYDADTGAULTNSFLETKAGSNQWYMGADGYAV-----1400
QY 1379 FYNLKGQLVGSGWETANHDWYIQSGKALTGEOTINGOHLIF-KEDGHQVKGQLVGTG 1437
Db 1401 -----RGHQTIOSRHMYFDAETGQQAQGIWVTD 1429
QY 1438 DGKVRYYDANSQDAFNKSVTVNGKTYVFGNDGTATAGNPKGQTFKDGSDIRFYSMEQ 1497
Db 1430 NGRKYPYDANTGDRVNVQVLVNGSWYFFGYQAA-----1464
QY 1498 LVTGSGWYENAOQMLYVKNQKVLTLGLQTVGSRVYFDENGIQAKGKAVRTSDGKIRYVD 1557
Db 1465 -----VTGFRDIRGQHLIFYNPDGTGAQGTTKVI-DNRIYTFD 1500
QY 1558 ENSGSM 1563
Db 1501 ADSGEL 1506

RESULT 13
Q56CX8_9STRE

ID Q56CX8_9STRE PRELIMINARY; PRT; 1506 AA.
AC Q56CX8;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DE Glucosyltransferase-T.
GN Name=gtft;
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OX Streptococcus.
NCBI_TaxID=13110;
RN 1)
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=B13N;
RA Kuwahara N.;
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
KW EMBL; AY966490; AAX76986.1; -; Genomic_DNA.
SQ SEQUENCE 1506 AA; 167016 MW; 94B882BP2C17C451 CRC64;

Query Match 40.8%; Score 3410.5; DB 2; Length 1506;
Best Local Similarity 43.4%; Pred. No. 5e-149;
Matches 708; Conservative 277; Mismatches 446; Indels 197; Gaps 39;

QY 1 MEKNVRFKWHKVKKRWVTLVSASATMLASALGASVASADTDTASDD--SNQAVVTGDQTT 58
Db 1 MERKLHYKLHKVKQWVTVIAVASAG-LASIVGA--GSLQTVSADDLAKEQAAASQOKAA 57
QY 59 NNQATDQT-----SIAATATSE-----QSASTDAAT-----DOASAAEQTGTTASTDT 102
Db 58 ANQNEDEVASDAADTASAKATSEKEVVGSSDTNSETNQVETKQOASAKESADAVAKQAPQ 117
QY 103 AAQTTN---ANEAKWVFTENENQGTDEMLAEAKNVATAESDSIPSLAKMNVKQVDG 159
Db 118 AGPATTISQVASSSSSVAPSKB---ADK--AAAGSVSQNEBEAALS---LANIKKIDG 167
QY 160 KYIYDQDGNVKKNFAVSVGDKIYYPD-ETGAYKDTSKVDADKSSSAVSQNAITIAANNR 218
Db 168 KYIYVMDGSKYKKNPAITVDGQMLYFDKTAGLSSTSYSPSQGLTPIVSD---FSVNNK 224
QY 219 AYSTAKNFEAVDNYLTADSWVRPKSILKDGKWTPESGKDDFRPLLMAWMPDTEKRYNV 278
Db 225 AFDSEKSEFELVDGVLTAESWVRPAKILENGKTWVDSKETDLRPVLSMWPNKDTQVAYL 284
QY 279 NYMNVGIGDKTYTAETSQADLTAAAEVLQARIKQITSENNTKWLRAISAFVFTQPOW 338
Db 285 NYMSKALGKGEFTTETSQTLTAAEELIQAKIARVSKEQGTWLRREMAAFVATQSRW 344
QY 339 NGESEKPYD--DHLONGALLFDNQDTLTPDTQSNVRLNRTPTNQTGSLDSRFTYNPNDP 396
Db 345 NKDSEQ-YDKADHLQGGALLYTN--NNLTWANSNWRLLNRTPTRO----DGKTHYSKADK 398
QY 397 LGGYDFLLANDVDSNPVVOAEQLNWLHYLLNFGSIYANDADANFDSIRVDAVNDVDADL 456
Db 399 YGGTEFFLLANDVDSNPVVOAEMLNQIHYLLNMGELVMGDKNANFDGIRVDAVDVNDAT 458
QY 457 LQISDYLKAAAYGIDKNNKNANHVSIIVEAMSDNDPTPYLHDDGDMNMNDKFRSLMLS 516
Db 459 LQLYTNFNSVGVNKSQAALAHISVLEAWSYNDNDYNDQNTNGAALAMNDGLRFSLLY 518
QY 517 LAKPLDKES-GLNPLIHN--SLVDREVDREVEVTPSYSPFAHADSEVQDIIRDIKAEI 573
Db 519 LTRPNIETPGMSTLIKSEYGLTDRTKNDKYDQTPSYVFNRAHSEVQTVTAQIIKEKI 578
QY 574 NPNSFGYSFTOBEIRQAFKIYNEDLKTKYTHYNVPLSYTLTLTKNGSI PRVYVGMNF 633
Db 579 DPTTDTGFTLLDQLKQAFELYNKONSVNKHYYTHYNI PAAYAVMLSNMESVTRVYVYGLD 638
QY 634 TDDGQYMAKNTVNDYDAIESLLKARKMYVSGQAMO--NYQ-----IGNGE 676
Db 639 TDDGQYMAKSPYYDAINTLLRARIRYAAGGQIMEHNSYKPSAAMKAHPADAGNVLGSE 698
QY 677 ILTSVRYGKALKQSD-KGDATTRTSVGVVGNQPNFSLD-GKVVALNMGAHAHQEYR 734

Search completed: February 11, 2006, 19:38:29
Job time : 223.546 secs

589 QAPKINEDLKTDKXKTHYVPLSYLLTNKGSIPRVYVGMFTDDGOYMANKTWYD 648
669 QAFYINADELKADEYATYNI PASYAVLLTNKDTVPVYVGLDFSDGOYMSQKSPYD 728
649 AIESLLKARMKYVSGOAM-----QNYQINGE-----ILTSVRKGAQKQSDKGDAT 697
729 AITSLKSRKXYVAGQSMNMYLHCFDPKAKNETKPGQVLTSVRYGKAMTADDLGNSD 788
698 TRTSGVVMGNQPNFSL-DGKVVALNMGAAHAHQYRALMVSTKDG VATYATDADASKA 756
789 TROQIGLVLNNKPFLLNDDDEIVLNMGAHKNQAYRPLMLTTKSLGLIYDKDAGAP-- 846
757 GLVKTDENGYLYFLNDDLKGVANPQVSGFLQVWVPVGAADDQDIRVAASDTASTDGKSL 816
847 --VVTNDAGQLIFKSDMYVGSNPQVSGYFAAWVPVGASDSQDARTQSSQSETKGDVY 904
817 HODAAVMSVMEFSGFNSFQSFATKEEYTNVVIANNVDFVSWGIITDPEMAPOYVSSTDG 876
905 HSNALDSNVIYEGFSNFQAMPEKNDFTNVKIAQNAKLFKOLGITSFELAPQYRSSTDN 964
877 QFLDSVIQNGYAFTRDYDLGMSKANKYGTADOLVKAIKALHAKGLKVMADWVPDQMYTFP 936
965 SFLDSVIQNGYAFTRDYVGYNTPYKGTVDQLLSRLAHAGQIQAINNDWVPDQIYNLP 1024
937 KOEVTVTTRDKFGKPIAGSQINHSLYVTDTKSSGDDYQAKYGGAFDELKEKYPELFTK 996
1025 GEQIVTAVRTNGSGKYDYSVINNTLYDSRTVGG--EYQEKFGGLFLDQKXDYPSLPET 1083
997 KOISTQOATDPSVKIKQWAKYFNGSNILGRGADYVLSQVSNKYFNVASDTLFLPSSLJ 1056
1084 KOISTQNPMPVKIKWAKYFNGSNILGRGADYVLSQVSNKYFNVASDTLFLPSSLJ 1143
1057 KRWVBSG-IRYDGKGIYVNSSATGDOVKASFTEAGNLYYFGKGVMTGAOTINGANYF 1115
1144 GEKSTGFTENGKTSFYSTS--GYQAKDTFTQDGTNWTYFVFNAGYMLTGKQNIHDKNYI 1201
1116 FLENGTALNTIYTDAGQNSHYAND-GKRYENG-YQFQNDWRY-FKDGNNMAGLITV- 1171
1202 FLPNGVELQDAYLFD--GNQEFYVYKAGEQVMNQYQDSQNMWHYFFENGRAIGLITEVP 1259
1172 --DGN--VQYFDPKDGVOAKDKLIVTRDGKRVYFDQHNGNAATNTP--IADKTHWYLYGK 1225
1260 NADGTHVTQYFDANGVQIKGTAKQNNQLRYFDEATGNMVMVNSWGLADKS--WLYLNA 1317
1226 DGVAVTGAQTVCKQKLYPEANGQVKDPVTSDEKLYFYDVSQDMWTDTFIEDKAGNW 1285
1318 QGVATGNQKIDGEEYFNADGQVKGNAIIDNNGDQRYDGDGKGMVMVNSWGLPDSW 1377
1286 FYLGKDGAAVTGAQTRIGQKLYFKANGQVKGDI VKGTDGKIRYDAKSGEQVFNKTVKA 1345
1378 LYLNDKGIATGROVINNVNFFGNDGKQIKDAFKLLSDGSMVYLDKDG---LITTGAKV 1434
1346 ADGKTYVIGNDGVAVDPVSVKQGTQFKDASGALRFNLK-GQVLTGSGWYETANHDWYI- 1403
1435 INGLMFFDKDG-----HQIKGDASTDANGKRYDYDKNDGHLVTNS-WGELPDGSLYLE 1488
1404 QSGKALTCBOTINGOHLFKEDGHQVKGOLVTGTG-----KURY 1443
1489 EQGDVAVTGQVJDGKTRFYDEDEGKQIKNSLKTFLANGDKIYLDGDGVAATGLQHVGDKIMY 1548
1444 YDANSQDAFNKSVTV-NGKTYVFNNDGTAQTA-GNPKQTF-----KDGSDIR----- 1490
1549 FD-EDGKVQVGVKFSVAKDGSWYLLNQDQVAAVGPSSINGQSLYFDQDGKQVKNVNRNSD 1607
1491 -----FYSMEGQVLTGSGWYENAOQWLYV-KNGKVLTLGLQTVGSGQRYVFDENGIOAKGK 1544
1608 GTTNYTGLTGEKLT-QDFGELPDGSWIYLDQAQGHVTVGAQIINGQNLVFKADGQVKGH 1666
1545 AVRTSDGKIRYFDENSGSMITNQWKFVY-QYVYFNDGARI 1585
1667 AYTDLGHMRFYDPSDGMLSNRFEQITFGVWAYFGADGVAI 1708

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 18:56:56 ; Search time 154.681 Seconds
(without alignments)
4414.224 Million cell updates/sec

Title: US-10-797-821-38
Perfect score: 8237
Sequence: 1 MEKKLHYLKHVKHWTIA.....FVRIGDQWTFNGDGAATNL 1554

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 244163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*
9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8237	100.0	1554	7	ADD93658 Streptoco
2	8237	100.0	1554	9	ADx37281 Streptoco
3	3818	46.4	1518	7	ADD93660 Streptoco
4	3818	46.4	1518	9	ADx37283 Streptoco
5	3615	43.9	1430	5	Aau98029 S. mutans
6	3615	43.9	1430	7	ADD93656 Streptoco
7	3615	43.9	1430	9	ADx37279 Streptoco
8	3610	43.8	1430	5	Aau98043 S. mutans
9	3609	43.8	1430	5	Aau98042 S. mutans
10	3609	43.8	1430	5	Aau98041 S. mutans
11	3604	43.8	1430	5	Aau98044 S. mutans
12	3604	43.8	1430	5	Aau98045 S. mutans
13	3502.5	42.5	1590	9	ADD93657 Streptoco
14	3502.5	42.5	1590	7	ADx37280 Streptoco
15	3497	42.5	1476	5	Aau98041 S. mutans
16	3472.5	42.2	1592	2	AAR32925 Glucosylt
17	3446.5	41.8	1475	5	Aau98036 S. mutans
18	3445.5	41.8	1475	5	Aau98035 S. mutans
19	3445.5	41.8	1475	5	Aau98034 S. mutans
20	3442.5	41.8	1475	5	Aau98038 S. mutans
21	3440.5	41.8	1475	5	Aau98037 S. mutans
22	3440.5	41.8	1475	5	Aau98032 S. mutans
23	3439.5	41.8	1475	5	Aau98031 S. mutans
24	3435.5	41.7	1475	5	Aau98030 S. mutans

25	3434.5	41.7	1475	5	Aau98027 S. mutans
26	3434.5	41.7	1475	7	ADD93654 Streptoco
27	3434.5	41.7	1475	9	ADx37277 Streptoco
28	3430.5	41.6	1475	5	Aau98040 S. mutans
29	3429.5	41.6	1577	2	AAR91047 Alpha-D-g
30	3428.5	41.6	1475	5	Aau98033 S. mutans
31	3407.5	41.4	1475	5	Aau98039 S. mutans
32	3255	39.5	1499	7	ADCS4806 Protein S
33	3161.5	38.4	1375	5	Aau98028 S. mutans
34	3161.5	38.4	1375	5	Aau98028 S. mutans
35	3161.5	38.4	1375	7	ADD93655 Streptoco
36	3161.5	38.4	1375	9	ADx37278 Streptoco
37	3142	38.1	2835	5	ABR98574 Dextran S
38	3142	38.1	2835	6	ABR55594 Amino aci
39	3080.5	37.4	1365	7	ADD93659 Streptoco
40	3080.5	37.4	1365	9	ADx37282 Streptoco
41	3041	36.9	1527	7	ADCS4807 Leuconost
42	3015	36.6	1527	5	Aau80055 Leuconost
43	2958	35.9	1477	9	ADY72696 Mutant de
44	2953	35.9	1477	9	ADY72733 Mutant de
45	2948	35.8	1477	9	ADY72732 Mutant de

ALIGNMENTS

RESULT 1
ADD93658
ID ADD93658 standard; protein; 1554 AA.
XX
AC ADD93658;
XX
DT 29-JAN-2004 (first entry)
XX
DE Streptococcus sobrinus glucosyltransferase-U.
XX
KW Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.
XX
OS Streptococcus sobrinus.
XX
PN WO2003075845-A2.
XX
PD 18-SEP-2003.
XX
PF 07-MAR-2003; 2003WO-US006962.
XX
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
XX (FORS-) FORSYTH INST.
XX Smith DJ, Taubman MA;
XX WPI; 2003-845091/78.

Composition useful as vaccines for dental caries comprises a fragment of a glucan binding protein-B binding to a major histocompatibility complex class II protein.
Claim 16; Page 15; 49pp; English.

The present sequence is the protein sequence of Streptococcus sobrinus glucosyltransferase-U (GTF-U). Peptide fragments of GTF-U, especially from the catalytic domain of the polypeptide, can be used in immunogenic compositions and subunit vaccines for dental caries. These compositions comprise a major histocompatibility complex (MHC) class II protein-binding peptide from S. mutans Glucan binding protein-B (GbpB) covalently linked with a peptide fragment of a streptococcal glucosyltransferase. The compositions are used in a claimed method of eliciting production of an antibody in a mammal. Dieptopic or multiepitopic polypeptides can be prepared synthetically or by recombinant DNA technology. Antibodies raised against MHC class II binding fragments of GbpB can be used in passive immunisation.

XX New composition comprising a fragment of a glucan binding protein-B
PT (GbpB) that binds to MHC class II protein, and a biocompatible
PT microparticle, useful for producing an antibody (claimed) for immunizing
PT mammals against dental caries.

XX Claim 7; SEQ ID NO 38; 73pp; English.

XX The invention relates to a composition comprising a fragment of a glucan
CC binding protein-B (GbpB) and a biocompatible microparticle, where the
CC fragment binds to a major histocompatibility complex (MHC) class II
CC protein. The composition is useful for producing an antibody for
CC immunizing mammals against dental caries. This sequence corresponds to a
CC Streptococcus sobrinus GbpB protein of the invention.

XX Sequence 1554 AA;

Query Match 100.0%; Score 8237; DB 9; Length 1554;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1554; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEKKLHYKLHKVKKHWVTIAVASIGLVSLVGAGTVAEDKVAANDTTAQATVGVDTGQDOA 60

DB 1 MEKKLHYKLHKVKKHWVTIAVASIGLVSLVGAGTVAEDKVAANDTTAQATVGVDTGQDOA 60

QY 61 TTNDANTNTTDTTADQSANNTQDQAGSDQSNQDQAKQDTANTDRNQADNSQTDNNQAT 120

DB 61 TTNDANTNTTDTTADQSANNTQDQAGSDQSNQDQAKQDTANTDRNQADNSQTDNNQAT 120

QY 121 DOATSPATDGTSVORRDAANVATAADQEGQTAPSEQEKSAALSLDNVKLIDGKYTVQAD 180

DB 121 DOATSPATDGTSVORRDAANVATAADQEGQTAPSEQEKSAALSLDNVKLIDGKYTVQAD 180

QY 181 GSYKKNFALTIVNGQMLYFSDTGALSSSTYSFSQGTTLNVDDFSSHAKAYDSTAKSPFL 240

DB 181 GSYKKNFALTIVNGQMLYFSDTGALSSSTYSFSQGTTLNVDDFSSHAKAYDSTAKSPFL 240

QY 241 VNGYLTANSWYRPAQILRNGQTWEASNENDLRPLMSWMPDKDTQVAYVYNNKYL SAN E 300

DB 241 VNGYLTANSWYRPAQILRNGQTWEASNENDLRPLMSWMPDKDTQVAYVYNNKYL SAN E 300

QY 301 TEVTNETSQVDLKEAQSQTQKI EQKITSDNSTQWLRTAMEAFVAAQPKWNMSTENFNKG 360

DB 301 TEVTNETSQVDLKEAQSQTQKI EQKITSDNSTQWLRTAMEAFVAAQPKWNMSTENFNKG 360

QY 361 DHLQGGALLYTNSDLTPWANSYRLLNRPQTQDGTCKYFTGEGGVEFLLSNVDNSN 420

DB 361 DHLQGGALLYTNSDLTPWANSYRLLNRPQTQDGTCKYFTGEGGVEFLLSNVDNSN 420

QY 421 PVVQAEQLNQLHYLMNWGDI VNGDKDANFPGVVRVDAVDNVDNADLLQVYSNYFKDNYKVTD 480

DB 421 PVVQAEQLNQLHYLMNWGDI VNGDKDANFPGVVRVDAVDNVDNADLLQVYSNYFKDNYKVTD 480

QY 481 SEANALAHISILEAWSLNDQYNEDTNGTALSIDSNSRLTSLAVITKQPGQRIDLSNLIS 540

DB 481 SEANALAHISILEAWSLNDQYNEDTNGTALSIDSNSRLTSLAVITKQPGQRIDLSNLIS 540

QY 541 ESNVNERANDTAYGDTIPTYSFVRADHSEVQTVIAKIVKEKIDTNSDGYTFTLDQDKAF 600

DB 541 ESNVNERANDTAYGDTIPTYSFVRADHSEVQTVIAKIVKEKIDTNSDGYTFTLDQDKAF 600

QY 601 KIYNEDMAKVNKTYTHYIPAAAYALLSNMESVPRVYVYGDLYTDDQYMAKSPYDAIA 660

DB 601 KIYNEDMAKVNKTYTHYIPAAAYALLSNMESVPRVYVYGDLYTDDQYMAKSPYDAIA 660

QY 661 TMLQGRIRIAYVSGQSEEVHKVNGNNOILSSVRYGQDLMSADTQGTDLTSLRTSLGLTVLSN 720

DB 661 TMLQGRIRIAYVSGQSEEVHKVNGNNOILSSVRYGQDLMSADTQGTDLTSLRTSLGLTVLSN 720

QY 721 DPNLDLGGDSLTVNMGRAHANQAYRPLILGTFKDGQVQSYLKSDTNI VIKYTDANGNLTTFTA 780

DB 721 DPNLDLGGDSLTVNMGRAHANQAYRPLILGTFKDGQVQSYLKSDTNI VIKYTDANGNLTTFTA 780

QY 781 DDIKGYSTVDMGSLAVVVPVGAKGQDVRVAADTTNQKADGKSLKTSAAALDSQVYIEGFS 840

DB 781 DDIKGYSTVDMGSLAVVVPVGAKGQDVRVAADTTNQKADGKSLKTSAAALDSQVYIEGFS 840

QY 841 NFQDFANNADADYTNKKIAENADFFKKLGITTSFEMAPQYVSATDGSFLDSIIQNGYAFSDR 900

DB 841 NFQDFANNADADYTNKKIAENADFFKKLGITTSFEMAPQYVSATDGSFLDSIIQNGYAFSDR 900

QY 901 YDLAMSNNKNGYSKDDLALNALKALHANGIOAIADWVPDQIYQYLPGBEVVTAKTNSGNP 960

DB 901 YDLAMSNNKNGYSKDDLALNALKALHANGIOAIADWVPDQIYQYLPGBEVVTAKTNSGNP 960

QY 961 TFDAYINNALYATNTKSSGSDYQAOYGGAFDLDELKAKYPMFTVMNISTGKPIDPSTKIK 1020

DB 961 TFDAYINNALYATNTKSSGSDYQAOYGGAFDLDELKAKYPMFTVMNISTGKPIDPSTKIK 1020

QY 1021 QWEAKYFNGTNLVKGAGYVLSDDATGKYFTVNENGDFLPASTGDNQAKTGFYDGTGM 1080

DB 1021 QWEAKYFNGTNLVKGAGYVLSDDATGKYFTVNENGDFLPASTGDNQAKTGFYDGTGM 1080

QY 1081 AYSTSGNKAVNSFIYEGGHYYFYFDKGHMVTGSKYKAEDGNDYYFLPNGIQMRDAIYQDA 1140

DB 1081 AYSTSGNKAVNSFIYEGGHYYFYFDKGHMVTGSKYKAEDGNDYYFLPNGIQMRDAIYQDA 1140

QY 1141 QGNSYYRTGLYKGDNMWYPFVDPNNANKTVFHYFDANNVMAIGYRNMYGQYVYFDENG 1200

DB 1141 QGNSYYRTGLYKGDNMWYPFVDPNNANKTVFHYFDANNVMAIGYRNMYGQYVYFDENG 1200

QY 1201 FOAKGQLLTDDKGTHTYFDEDDNGAMAKKFNVDGDDWYMDGNGNAVKGQYVNNQILYFN 1260

DB 1201 FOAKGQLLTDDKGTHTYFDEDDNGAMAKKFNVDGDDWYMDGNGNAVKGQYVNNQILYFN 1260

QY 1261 PETGVQVKQFITDAQGRTSYYDANSYGALSSGFFTPNGSDWYAEVYVYKGFQVAEN 1320

DB 1261 PETGVQVKQFITDAQGRTSYYDANSYGALSSGFFTPNGSDWYAEVYVYKGFQVAEN 1320

QY 1321 QDQWYFDQTTQKQAKGAAKVDGRDLVFNPDGSGVQKGFATDESNTSFYHGDNDGKV 1380

DB 1321 QDQWYFDQTTQKQAKGAAKVDGRDLVFNPDGSGVQKGFATDESNTSFYHGDNDGKV 1380

QY 1381 GGFPTTGNNAWTYADNNGNLVKGFBIDGKWTHFDEVTQQQAKGAALVNGQQLYFVDVSG 1440

DB 1381 GGFPTTGNNAWTYADNNGNLVKGFBIDGKWTHFDEVTQQQAKGAALVNGQQLYFVDVSG 1440

QY 1441 IOVKGDFVTDGQNTSYDVNSGDKKVGFFTTGDNWYADGQGNLAKGRKSIDNQDLY 1500

DB 1441 IOVKGDFVTDGQNTSYDVNSGDKKVGFFTTGDNWYADGQGNLAKGRKSIDNQDLY 1500

QY 1501 FDPATGKQVKQGLVSDIGRNYFYDSSGNNMAKNRFVIRIGDQWYIFGNDGAAATNL 1554

DB 1501 FDPATGKQVKQGLVSDIGRNYFYDSSGNNMAKNRFVIRIGDQWYIFGNDGAAATNL 1554

RESULT 3

ADD93660

ID ADD93660 standard; protein; 1518 AA.

XX AC ADD93660;

XX DT 29-JAN-2004 (first entry)

XX ST Streptococcus salivarius glucosyltransferase-I.

XX KW Glucosyltransferase; enzyme; vaccine; antiserum; epitope; immunogen.

XX OS Streptococcus salivarius.

XX PN WO2003075845-A2.

XX PD 18-SEP-2003.

XX PF 07-MAR-2003; 2003WO-US006962.

XX XX

PR 07-MAR-2002; 2002US-0362209P.
PR 08-AUG-2002; 2002US-0402483P.
XX (FORS-) FORSYTH INST.
XX Smith DJ, Taubman WA;
XX WPI; 2003-845091/78.
XX
PT Composition useful as vaccines for dental caries comprises a fragment of
PT a glucan binding protein-B binding to a major histocompatibility complex
PT class II protein.
XX
XX Claim 16; Page 16; 49pp; English.
XX
XX The present sequence is the protein sequence of *Streptococcus salivarius*
CC glucosyltransferase-I (GTF-I). Peptide fragments of GTF-I, especially
CC from the catalytic domain of the polypeptide, can be used in immunogenic
CC compositions and subunit vaccines for dental caries. These compositions
CC comprise a major histocompatibility complex (MHC) class II protein-
CC binding peptide from *S. mutans* glucan binding protein-B (GbpB)
CC covalently linked with a peptide fragment of a streptococcal
CC glucosyltransferase. The compositions are used in a claimed method of
CC eliciting production of an antibody in a mammal. Diepitopic or
CC multi-epitopic polypeptides can be prepared synthetically or by
CC recombinant DNA technology. Antibodies raised against MHC class II
CC binding fragments of GbpB can be used in passive immunisation.
XX
XX Sequence 1518 AA;

Query Match 46.4%; Score 3818; DB 7; Length 1518;
Best Local Similarity 50.1%; Pred. No. 1.6e-193;
Matches 783; Conservative 215; Mismatch 434; Indels 130; Gaps 24;

QY 1 MEKLLHYKLHKYKHWVTIAVASIGLVLVG-----AGTVSA---EDKVANDTTAAQTAVG 52
DB 1 MENKLIHYKLHKYKHWVTIAVASVALATVIGLSVTTSSVSADETDQKTVTOSNGSTTAS 60

QY 53 VDTGDOQATNTDANTNTDADQSANINQDQAGSDOSNNQDQAKQDPTANTDRNOADS 112
DB 61 LVTSPEATKADKRTNKADVLTPAKETNAVETAT-TTNTQATAEAATTATTADVA-VA 118

QY 113 QTDNNQATQATSPATDGTSGVORRA---ANVATAADQEGQTAPSEQKSAALSLDNVKL 169
DB 119 AVPENKEAVVTTDAPAVTTEKAEQPAATVKAQVNTVEKAPAEALAKDSEVEAALSUKTN 178

QY 170 IDGKYVYVQADGSYKKNPAITVNGQMLYFDSDTGALSTSTYSFSOGTTLNVDDFSHNK 229
DB 179 IDGKYVYVNEQDGHENFAITVNGQLLYPGKD-GALTSSSTYSFTPGTNIYVDGFSINNR 237

QY 230 AYDSTAKSFELVNGYLTANSWRPAGILRNGQTWEASNENDLRPVLMSWMPDKDTQVAVY 289
DB 238 AYDSSEASPELIDGVLTDADSWYRPASIIKDGVTWQASTAEDFRLPMLMAMPNVVDVQVYL 297

QY 290 NYMKVLSANEVTEVNTSVDNLKEAQSIQTKIEQKITSNDSTOWLRTEAMEAFVAAQPK 349
DB 298 NYMKSVFNL-DAKYSSDQKQETLKVAAKDQIKIEQKIQAEKSTQWLRETFISAFKTPQ 356

QY 350 WNMSTENFKG---DHLQGGALLYTNSDLTPWANSDYRLNLTPTQDDCT--KKYFTEGG 404
DB 357 WNKETENS KGGEDHLQGGALLYNDSTPTWANSQYRLNLTATNTGTIDKSLDEQS 416

QY 405 E-----GGYEFLLSNVDNSNPVQAEQLNQLHLMNMGDIVMGDKDANFDGVRVADNVV 460
DB 417 DPNHMGDFDILLANDVLSNPVQAEQLNQIHYLMNMGSIVMGDKDANFDGIRVADNVV 476

QY 461 NADLLQVNSYFKDKVYVTDSEANALAHISILEASLNDNQYNEEDTNGTALSIDNSSRLT 520
DB 477 DADMLQLYNTYFREYVGNKSEANALAHISVLEAWSLNDNHYNDKTDGAALAMENKQRLA 536

QY 521 SLAVLTK-----PQQRIDLSNLSISESVNKERAND-----T 551
DB 537 LLFSLAKPIKERTPAVSPLYNNTFTTQREKTDWINKDGSKAYNEEDGTIVKOSTIGKYNE 596

QY 552 AYGDITPTYSFVRAHDSVQTVIAKIVKEKIDTNSDGYTFLDQLDKAFKINVEDMAKN 611
DB 597 KYGDASGNVVFIRAHNNVQDIIAIIKKEINPKSDGFTITDAEMKQFAEINKDMLSSD 656

QY 612 KTYTHYNIPAAVALLLSNMESVPRVYGDLYTDDQYMAKKSPPYDAIATMLQGRYAVYS 671
DB 657 KKYTLNIPAAVAVMLQNMETITRVYGDLYTDDGHYMETKSPYVDITVNLKMSIKYVS 716

QY 672 GQSQSEVH-----KVNGNNQILSSVRYGQDLMASADDTQGTDLSTGSLVTLV 718
DB 717 GQQAQRSVWLPDGTGMDNSDVELYRTNEVTSVRYKDIIMTANDTEGSKYSYTSQVTLV 776

QY 719 SNDPNLDLGGDS-LTVNMGRAHQAQYRPLILGTGQVQSYLKDSD---TNIVKYTDANG 774
DB 777 ANNPNLNDQAKLNVEMGKIHANQYRALIVGTADGINKFTSDADAJAAGYVKETDSNG 836

QY 775 NLFTTADDIKGVSTVDMSGYLAVVVPVCAKQDQVRVAADTNOKADGK-SLKTSAAALDSQ 833
DB 837 VLTFGANDIKGYETFDMSGFVAVVVPVQASNDQIRVAPSTPAKKEGELTLKATEYDSQ 896

QY 834 VIYEGFSNFQDF--ANNADADYTNKKAENADFFKKLGITSPEMAPQYVSATDGSFLDSII 891
DB 897 LIYEGFSNFQITPDGSDPSVVTNRKIAENVLDLFXSGVGTSEMAPQYVSADDTGFLDSVI 956

QY 892 QNGYAFSDRYDLAMSKNNKYGSKDDLALNALKALHANGIOATADWPDQIYQLPGSEVVT 951
DB 957 QNGYAFADRYDLAMSKNNKYGSKEDLRDALKALHKAGIQAIADWVPDQIYQLPGSEVVT 1016

QY 952 KRTSYNGNPTFDAYINNALYATNTKSGSDVQAOYGGAFDLDELKAKYPMDFVNNMISTGK 1011
DB 1017 TRTDGAGKIDAIIDHSLYVANSKSSKDKYQAKYGGFEFLAKKAKYPMDFVNNMISTGK 1076

QY 1012 PIDPSTKIKQWEAKYFNGTIVLKGAGYVLSDDATGKYFTVNNENGDFLPPASFTGPNQAKT 1071
DB 1077 PIDDSVKLKQWKAERYFNGTIVLGRGVYVLSDEATGKYFTVTKEGNFIPQLTGTKEKVT 1136

QY 1072 GFYDGTQMAYYTSGKAVNSFIYEGHYHYFDKGHWYTSYKABDNDYY-FLPNGI 1130
DB 1137 GFSSDGKGIYFGTSGTQAKSAFVTFNGTNYFFDARGHMTNSEYSPNGKQVYRFLPNGI 1196

QY 1131 QMRDAIYODAGNSYVYGRGTGILYKGDNWPVDPDPNNANK-----TVFRYFDANNVMA 1183
DB 1197 MLSNAFYIDANGNTYLYNSKQMYKGG--YTKFDVSETDKDKESKVVVKFRYFTNEGVA 1254

QY 1184 IGYRNMVQTYFDBENGFAQQLTDDKGYHFDENGDGAMAKNFVNVGDDWYMDGNG 1243
DB 1255 KGVTVIDGFTQYFGEDGFOAKDLVTFKGTVYFPAHTGNGIKDTRNINQKMYVFDANG 1314

QY 1244 NAVKQYVNNQILLYFNPETGVQVKQFITDAQGRTSYYDANSALKSSGFFTPNGSDWY 1303
DB 1315 VAATGAQVINGQKLYFN-EDGSQVKGGVVKNADGTYSKYKGFGBLVTNEFFTTDGNVWY 1373

QY 1304 YAEVGYVYKFKQVAENQDQWYYPDQTTGQAKGAQKVDGRDLYFNPDSGVQVKGDFATD 1363
DB 1374 YA-----GANGKTVTGAQVINGQKLYFNAD-GSQVKGGVVKN 1409

QY 1364 ESGNTSFYHGDNGDKVVGGFFTTGNNAWYADNNGNLVKGFQEIIDGKRWYHFDVDTGQOAK 1423
DB 1410 ADGTVSKYNASTGERLTNEFFTTGNNWYIYGANGKSVTGEVKI-----1453

QY 1424 GAALVNGQQLFPDVDSGIQVKGDFVTDGQGNSTYYDVNSGDKKYNNGFTTGTGNANWYADG 1483
DB 1454 -----GDDTYFFAKDGKQVKGQTVSAGNRIISYYIGDSGKRAVSTWIBIQPGVYVYFDK 1507

QY 1484 QG 1485
DB 1508 NG 1509

RESULT 4
ADX37283
ID ADX37283 standard; protein; 1518 AA.

XX AC ADX37283;
 XX DT 21-APR-2005 (first entry)
 XX DE Streptococcus salivarius glucan binding protein B.
 XX KW immunogenicity; immune stimulation; glucan binding protein-B;
 KW microparticle; major histocompatibility complex; tooth disease.
 XX OS Streptococcus salivarius.
 XX PN US2005031633-A1.
 XX PD 10-FEB-2005.
 XX PF 09-MAR-2004; 2004US-00797821.
 XX PR 13-APR-1998; 98US-0081550P.
 PR 08-JAN-1999; 199US-0115142P.
 PR 12-APR-1999; 99US-00290049.
 PR 07-MAR-2002; 2002US-0363209P.
 PR 08-AUG-2002; 2002US-0402483P.
 PR 07-MAR-2003; 2003US-00383930.
 XX (SMIT/) SMITH D J.
 PA (TAUB/) TAUBMAN M A.
 XX Smith DJ, Taubman MA;
 PI WPI; 2005-151644/16.
 XX
 DR New composition comprising a fragment of a glucan binding protein-B
 PT (GbpB) that binds to MHC class II protein, and a biocompatible
 PT microparticle, useful for producing an antibody (claimed) for immunizing
 PT mammals against dental caries.
 XX Claim 7; SEQ ID NO 40; 73pp; English.
 PS
 SS The invention relates to a composition comprising a fragment of a glucan
 CC binding protein-B (GbpB) and a biocompatible microparticle, where the
 CC fragment binds to a major histocompatibility complex (MHC) class II
 CC protein. The composition is useful for producing an antibody for
 CC immunizing mammals against dental caries. This sequence corresponds to a
 CC Streptococcus salivarius GbpB protein of the invention.
 CC
 SQ Sequence 1518 AA;
 Query Match 46.4%; Score 3818; DB 9; Length 1518;
 Best Local Similarity 50.1%; Pred. No. 1.6e-193;
 Matches 783; Conservative 215; Mismatches 434; Indels 130; Gaps 24;
 QY 1 MEKKLHYKLHKVKKHWTIAVASIGLVSLVG-----AGTVSA---EDKVANDTTAAQTVG 52
 DB 1 MENKTHYKLHKVKKQWTTIAVASALATVLGSLVTTSSVSADETQDKTQVTSQSGTTAS 60
 QY 53 VDTGQDQATNTDANTNTDADQSANNTQDQAGSDQSNODQAKQDTANTDRQADNS 112
 DB 61 LVTSPEATEAKRINTKEADVLTTPAKETNAVETAT-TTNTQATAEATATTADVA-VA 118
 QY 113 QTDNNQATDQATSPATDGTSGVRRDA---ANVATAADQEGOTAPSEQEKSAALSLDNVKL 169
 DB 119 AVPNKEAVVTTDAPATTEKAEQPAQVKAEEVNTVEKPAEALKDXSEVAALSLSKNKN 178
 QY 170 IDGKYYVQADGSYKKNFAITVNGQLYDSDTGLSSTSTVSFSGQTTNLVDDFSSHNK 229
 DB 179 IDGKYYVNEGDGSHKENFAITVNGQLYFGKD-GALTSSSTVSFTPGTTNI VDGFSINN 237
 QY 230 AYDSTAKSFELVNGYLTANSVRRPAGILRNQGTWEASNENDLRPVLMSSWPKDQTVAYV 289
 DB 238 AYDSSEASFELIDGYLTADSWTPASIIKDGVTWQATDPRPLLMAMWPNVQVNYL 297
 QY 290 NYMKNYLSANETEVNTNETSQVDLNKEAQSIQTKIEQKITSNDSTQWLRTAMEAFVAAQPK 349

DB 298 NYMSKVFNL-DAKYSTSDKQETLKVAAKOIQIKIEQIOAEKSTQWLREITSAFVKTPQ 356
 QY 350 WNMSTENFNKG---DHLQGGALLYTNSDLTPWANSYRLNLRPTQDQGT--KKYFTTEGG 404
 DB 357 WNKETENYKGGGEDHLOGGALLYVNDSTRTPWANSYRLNLRPTATNTQGTIDKSLDEQS 416
 QY 405 E-----GGYBFLLSNDVDNSNPVVAQBLQNLHYLMWGDIVMGDKDANFGVRVDAVDN 460
 DB 417 DPNHMGGFDELLANDVDLSNPVVAQBLQNLHYLMWGSIVMGDKDANFGIRVDAVDN 476
 QY 461 NADLLQVSNYKONYKVTDSANALAHISILSAWSLNDNQYNEEDTNGTALSIDNSSRLT 520
 DB 477 DADMLQLTYNYPREYGVNKSSEANALAHISVLEAWSLNDNHYNDKTDGAALAMENKQRLA 536
 QY 521 SLAVLTK-----QPGQRIDLSNLISESVNKERAND-----T 551
 DB 537 LLFSLAKPIKERTPAVSPLYNNTNTTQDEKTDWINKDGSKAYNEDGTVKQSTIGKYNE 596
 QY 552 AYGDITPTYSFVRAHDSVQTVIAKIVKEKIDTNSDGYTFTLDLQKDAFKIYNEDMAKN 611
 DB 597 KYGDASGNTVFIHANNVQDIIAEIIKEINPKSDGFTITDAEMKQAEFIYNKMLSSD 656
 QY 612 KTYTHYNIPAAVALLSNMESVPRVYVYGDLYTDDGQYMAKSPYDAIATMLQGRYAVS 671
 DB 657 KKYTLNNIPAAVAVMLQNMETITRVYVYGDLYTDDGHYMETKSPYDITVNLMSRIKYS 716
 QY 672 GQGSSEVH-----KVGNNQILSSVRYQDILMSADDTQGTDLRSRTSGLVTLV 718
 DB 717 GQGAQRSYWLPTDGMKDNDSVELYRTNEVYTSVRYGKIDMTANDTEGSKYSRTSGQVTLV 776
 QY 719 SNDPNLDLGGDS-LTVNMGRAHANOAYRPLILGTGQVOSYLKQSD---TNIVKYTDANG 774
 DB 777 ANPNKLNLDOSAKLVEMGIHANOKYRALIVGTADGIKNFTSDADAIAAGYVKETDSNG 836
 QY 775 NLTFTADDIKGYSTVDMSGYLAVVVPVGAQDQVRVAADTNTQKADGK-SLKTSAALDSQ 833
 DB 837 VLTFGANDIKGYETFDMSGFVAVVVPVGAQDQIRVAPSTEAKKEGELTLKATEAYDSQ 896
 QY 834 VIYEGFSNPDF--ANNDDATNKKIAENADFPKLGITSPFEMAPQVYVATGSPFLDSII 891
 DB 897 LIYEGFSNFQTTIPDGSDFSVYTNRIKENVDLFSKMGVTSFEMAPQVFSADDTFLDSVI 956
 QY 892 QNGYAFSDRYDLAMSKNNKYGSKDDLANALKAHANGIOAIADWPDQIYQVLPGERVITA 951
 DB 957 QNGYAFADRYDLAMSKNNKYGSKEDRLKALHAKAGIOAIADWPDQIYQVLPGERVITA 1016
 QY 952 KRTSYGNPTFDAYINNALYATNTKSSGSDYQAQYGGAFDELKAKYPMFTVNMISTGK 1011
 DB 1017 TRTDGAGRKIADAIIDHSIYVANSKSGKDYOAKYGGEFLEALKAKYPMFKVMISTGK 1076
 QY 1012 PIDPSTKI KQWEAKYFNGTNVLGKAGYVLSDDATGKYFTVNVNENGDFLPASFTGQNAKT 1071
 DB 1077 PIDDSVKLQWKAERYFNGTNVLERGVYVLSDEATGKYFTVTKEGNFIPLQLTGKEKVI 1136
 QY 1072 GPYVDGTGMAYYSTSGNKAVNSFIYEGGHYVYFDGHWVTGYSYKAEQDNDYY-FLPNCI 1130
 DB 1137 GFSSDGKGITFTGSGTQAKSAFVTFNGNTYTFDARGHMTVNSEYSPNGKDYVRFPLNGI 1196
 QY 1131 QMRDAIYODAQNSYYYGRTGILYKGDNMWPPVDPNNANK-----TVFRYFDANNVMA 1183
 DB 1197 MLSNAFYIDANGNTLYNSKGMQYKGG--YTKFDVSETDKDGESKVKYKFRFTNEGVA 1254
 QY 1184 IGRNMYGQTYFDBNGFOAKGQLTDDKGTHTYFDENGMAMAKFVNVVDDWYMDGNG 1243
 DB 1255 KGVTVIDGFTQYFGBDGPQAKDKLVTFKGTYYFDAHTGNGIKDTRNKGWYFDFANG 1314
 QY 1244 NAVKQYVNNQILYFNBEETGVQKQFITAQGRTSYVDANSGLKSSGFFTPNGSDWY 1303
 DB 1315 VAATGAQVINGOKLYFN-EDGSQVKGKGVVKNADGTYSKYKEGFGELVTNEFFTTDGNVY 1373
 QY 1304 YAEANGYVYKGFQKQVAENQDQWYFYDQTTGKQAKGAQKDRDLVFNPDPSGVQVKGDFATD 1363

Db 532 LVALTR-FLEK-DASNKNEIRSGLEPVITNSLN-NRSAEGKNSRMANYIFIRAHDSVQ 588
Qy 572 TVIAKIVKEKIDTNSDGYFTTLOQLKDAFKINEMAKVNKTYTHYNIPAAVALLSNME 631
Db 589 TVIAKIIKAQINPKTDGLTFTLDELKQAFKINEDMRQAKKYTQSNIFTAYALMLSNKD 648
Qy 632 SVPRVYGYDLYTDDGOYMAKKSPYDAIATMLQGRITAYVSGSBEVHKVNGNQ----- 686
Db 649 SITRLYGYDMYSDDGOYMAKSPYDAIDITLLKARIKYAAGGDMKITTYVEGDKSHMDWD 708
Qy 687 ---ILSSVRYGQDLMSADTQGTDLRSRTSLVTLVNSNDPNLDLG-GDSLTVNMGRAHAKQ 742
Db 709 YTGVLTSVRYGTGANEAATD-QQSEATKTQGMVITSNPSELKLNQNDKVIIVNMGAAHKQ 767
Qy 743 AYRPLILGTGQVQSYLKSDST-NIVKYTDANGNLFTTADDIKGYSTVDMSGYLAVWVPV 801
Db 768 EYRPELLLTTKDGLTSYTSDAAKSLYRKNTDKRGELVFDASDIQGYLNPQVSYLAVWVPV 827
Qy 802 GAKDGQVRAADTNQKADGKSLKTSAAALDSQVIYEGFSNFODFANNADADYTKKIAENA 861
Db 828 GASDNQDVRAASNAKANATQVYESSALDSQLIYEGFSNFQDFVTYKSDYTNKKIAQNV 887
Qy 862 DPFKGLGITSFEMAPQVYSATDGSFLDSIIQNGYAFSDRYDLAMSKNKKYSGKDDLANAL 921
Db 888 QLFKSGVTSFEMAPQVYSSEDSFLDSIIQNGYAFEDRYDLAMSKNKKYSGQDDMINAV 947
Qy 922 KALHANGIOAIADWPDIQYQLPGEVVTAKRTNSYGNFTPDAYINNALYATNTKSSGD 981
Db 948 KALHKSIGIIVADWPDIQYVNLPGKEVVTATRVNDYGEYRKDSEIKNTLYAANTKSNKD 1007
Qy 982 YQAOYGGAFDELKAKYPMFTVNMISTGKPTDPTKIKQWEAKYFNGTNVILKGAGYVL 1041
Db 1008 YQAOYGGAFSELAAKYPIFNRTQISNGKKIDPSEKITAMKAKYFNGTNILGRGVGYVL 1067
Qy 1042 SDDATKYFTVNEGDFLPASFTGQONAKTGFYDGTGMAYTSTSGNKAIVSFIYEG-GH 1100
Db 1068 KQNASDKYFELKGNQTYLPKQWT-NKEASTGFVNDGNGMTFTSTGYQAKNSFVQDAKN 1126
Qy 1101 YYYFDKDHMTVGSYKAEKDNDYFIFLNGIQMRDAIYDQAQNSYYIYGRITGILYKGDNMY 1160
Db 1127 WYFFDNNGHMVYGLQL-NGEVQYFLSNGVQLRESFLENADGSKNYFGHLGNRY-SNGYY 1184
Qy 1161 PFVDPNNAKTVFRFPDANNVMAIGYRANMYGQTYFPDENGFOAKGOLLTDDKG-THYFDE 1219
Db 1185 SF-----DNDKRWYFPDASGVMAVGLKTINGNTQVDPQDGYQVKGAWITGSDGKRYFDD 1239
Qy 1220 DNGMAKNKPVN-VGDDWYMDGNGNAVKGOYPVNNQILYFNPETGVQVKGQFIFIDAQGR 1278
Db 1240 GSGNMAVNRFPANDKNGDWYLLNSDGLALGVQVQTINGKTYFG-ODGKQIKGKIITD-NGK 1297
Qy 1279 TSYDANSALKSSGFFTPNGSDWYIYAENGYYVYKFKQVAENQDOWYFPDQTGKQAKGA 1338
Db 1298 LKYFLANSGLARNIFAT-----DSQNNWYIFG-SDGVAVTGS 1334
Qy 1339 AKVGDGLYFNPDSGVQVKGDFATDESQNTSFPHGNDGKVGVGFFPTGNNAWYADNNG 1398
Db 1335 QTIAGKGLYFASD-GRQVKGVSFT-YNGKVHYHADSGELVNRFEADKGNWYLLDSNG 1392
Qy 1399 NLVKGFQEIIDGKWHYHDFEVTGQAKGAALVNGQLYFVDVDSGLOVKGDFVTD 1450
Db 1393 EALTGSQRI-----NDQRVFF-TREGKQVKGVDAYD 1422

RESULT 7

ADX37279

ID ADX37279 standard; protein; 1430 AA.

XX ADX37279;

AC ADX37279;

XX 21-APR-2005 (first entry)

XX Streptococcus mutant glucan binding protein B variant #8.

DE

XX immunogenicity; immune stimulation; glucan binding protein-B;
KW microparticle; major histocompatibility complex; tooth disease.
XX Streptococcus mutans.
XX US2005031633-A1.
XX 10-FEB-2005.
XX 09-MAR-2004; 2004US-00797821.
XX 13-APR-1998; 98US-0081550P.
XX 08-JAN-1999; 99US-0115142P.
XX 12-APR-1999; 99US-00290049.
XX 07-MAR-2002; 2002US-0363209P.
XX 08-AUG-2002; 2002US-0402483P.
XX 07-MAR-2003; 2003US-00383930.
XX (SMIT/) SMITH D J.
XX (TAUB/) TAUBMAN M A.
XX Smith DJ, Taubman MA;
XX WPI; 2005-151644/16.
XX New composition comprising a fragment of a glucan binding protein-B
XX (GbpB) that binds to MHC class II protein, and a biocompatible
XX microparticle, useful for producing an antibody (claimed) for immunizing
XX mammals against dental caries.
XX Claim 7; SEQ ID NO 36; 73pp; English.
XX The invention relates to a composition comprising a fragment of a glucan
XX binding protein-B (GbpB) and a biocompatible microparticle, where the
XX fragment binds to a major histocompatibility complex (MHC) class II
XX protein. The composition is useful for producing an antibody for
XX immunizing mammals against dental caries. This sequence corresponds to a
XX Streptococcus mutans GbpB protein of the invention.
XX Sequence 1430 AA;

Query Match 43.9%; Score 3615; DB 9; Length 1430;

Best Local Similarity 49.9%; Pred. No. 9e-193; Mismatches 418; Indels 112; Gaps 32;
Matches 745; Conservative 217;

Qy 1 MEKHLHYKLHKYKHVWTIAVASIGLSVLGAGTVSAEDKVANDTTAQATVGVDTGQ--- 57
Db 1 METKERRYKMHKVKHWTIVAVAS-GLIITL--GTTTLGSSVSAETEQQTSDKVVTKQSD 56
Qy 58 DOATTNDANTNTTDDTADQSNANTNQDQAGSDQSNNOQAKQDT-----ANTDR 106
Db 57 DKAAESSQTDAPKT---KQAQTEQTAQS-QANVADTSTSTIKETPSQNTTQANSDD 111
Qy 107 NQADNSQTDNNOATQATSPATDGTSVQRDAANVATA-ADQEGOTAPSEQESKAAL-SL 164
Db 112 KTVNTKSEEAQTSERTKQSEEAQTASSQALTAQALTKQRTAAQENKPNVDLAAI 171
Qy 165 DNVKLIDGKYVYVQADGSGYKKNFALTIVNGQMLYFDSDTGALSSTSYFSQGTNTNLVDFF 224
Db 172 PNVKQIDGKYVYIGSDGQPKKNFALTIVNKNVLYFDKNTGALTDTTSQYQFKQGLTKLNNDY 231
Qy 225 SHNKAYDSTAKSFELVNGYLTANSWYRPAGILRRGQWAEANENDLRPLMSWPPDKOT 284
Db 232 TPNQIVNFENTSLTIDNYVTADSWYRPKOILKNGKTWTASSESDRLPLMSWPPDKOT 291
Qy 285 QVAYVYNNKYLKLANETETVNTSQTDLNKEAQSITKIEOKITSDNSTOWLRTAMEAFV 344
Db 292 QIAYLNTNQQGLGTGENVTADSSQESLNLAATQVQVLIETKISQTOQTQWLRLDINSFV 351
Qy 345 AAQPKNNMSTE---NFKRGDHLQGGALLYTNSDLTPWANSDYRLNRLTPTQODGTKKYFT 401
Db 352 KTQPNWNSQTESDTSAGEKDHQGGALLYSNSDKTAYANSDYRLNRLTPTSGTKPKYPE 411

CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
CC is produced in the amyloplast of potato or the vacuole of sugar beet.
CC Glucans are useful as substitutes for and additions to modified starch
CC and latexes in paper manufacture. Unlike prior art techniques, which
CC require input materials that produce chemical effluents, paper
CC manufacture utilising the glucan produced by GFP, which utilises
CC biologically produced input materials, is more cost-effective and
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
CC properties and impart gloss to the paper during coating step. The present
CC sequence represents a GFPD mutant of the invention. Note: The present
CC sequence is not shown in the specification but was created by the indexer
CC using the GFPD sequence appearing as AAU98029 and the information in
CC claim 36
XX
SQ Sequence 1430 AA;
Query Match 43.8%; Score 3610; DB 5; Length 1430;
Best Local Similarity 49.9%; Pred. No. 1.6e-182;
Matches 744; Conservative 218; Mismatches 418; Indels 112; Gaps 32;
QY 1 MEKKLHYKLVKXKHWVTIAVASIGLVSLVGAGTVAEDKAVANDTTAQTAVGVDTGQ--- 57
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
1 METKRYKMKHKXKHWVTIAVAS-GLITL---GTTTLGSSVSAETEQTSDKVVTQKSED 56
QY 58 DOATTNDANTNTDTTADQSANTNQDQAGSDQSNQDQAKQDT-----ANTDR 106
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
57 DKAAESSQTDAPKT-----KQAEQTQTAQS-QANVADTSTSIKETPSONITTOANSDD 111
QY 107 NQADNSQTNNOATDOATSPATDGTSVORRDAANVATA-ADQEGGOTAPSEQEKSAAL-SL 164
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
112 KTVYTKSEATSEBTQKSEATQASSQALTQAKAELTKQRTAQAKNPNVDLAAI 171
QY 165 DNVLKIDGKYVVOADGSKYKXNFATVNGOMLYFSDTGALSGSTSYFSQGTTLNVDVF 224
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
172 PNVKQIDGKYVIGSDGQPKNFALTANNKLVYFDKNTGALTDTSQYQKQGLTKLNNDY 231
QY 225 SHNKAYDSTAKFELVNGYLTANSYRPAAGILRNGQWTEASNENDLRPLMSWPDKDT 284
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
232 TPHNQIVNFENTSLETIDNYVTADSWYRPKDLKNGKTWTSSESDLRPLMSWPDKDT 291
QY 285 QVAYVYNNKYL SANETEVTNETSOVDLNKEAQSIQTKIEOKI TSDNSTQWLRTAMEAPV 344
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
292 QIAYLNNYMQGLGTGENYATDSSQESLNLAAQTQVVKIETKISQTOQTQWLRLDINSFV 351
QY 345 AAQPKXNMSTE--NPNKGDIHLOGGALLYTNSDLTPWANSYRLLNRTPTQDGTGKXYT 401
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
352 KTQPNWNSQTESDTSAGEKXHLQGGALLYSNSDKTAYANSYRLLNRTPTSGPKPYE 411
QY 402 EGEGEGYELLNDVNSNPVQAOLNOLHLYLMWNGDIVMGDKDANPDGVRVADVNVN 461
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
412 DNSSGGYDFLLANDIDNSNPVQAOLNOLHLYLMWNGDIVANDPEANPDGVRVADVNDV 471
QY 462 ADLLQVSYNYFKDNYKVTDSEANALAHISILEAWSLNDNQYNEDTNGTALSIDNSRLTS 521
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
472 ADLLQIASDYLKAHYGVDSKEXNANHLISILEASNDNPQYNKDTYKQALPIDNKLRLSL 531
QY 522 LAVLTKQPOQRIDLN-----LISESVNKERANDTAYGDTIPTVYSFVRAHDSEVQ 571
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
532 LYALTR-PLEK-DASNKNEIRSGLEPVTNSLN-NRSABGKNSRMANYIFIRAHDSVQ 588
QY 572 TVIAKIVKEIDTNSDGYFTLIDOLKDAFKIYNEDMAKVNKYTYTHYNIPAAVALLSNME 631
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
589 TVIAKIIIAQINPKTDGLFTLDELKQAFKIYNEDMRQAKKYQTSNIFTAYALMLSNKD 648
QY 632 SVPRVYVGLYTDGQYMAKSPYYDAIATMLQGRYAYVSGSGSEVHKVNGNQ----- 686
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
649 SITRLYVGYDDGQYMAKSPYYDAIDTLKARIKAYAGQDMKIITYVEGDKSHMDWD 708
QY 687 ---ILSSRVYQDILMSADTQGTDLISRTSGLVTLVSNDPNLDLG-CDSLTVNNGRAHQ 742
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
709 YTGVLTSVRVYGTGANEATD-QGSEATKTQGMVITSNPSPSLKLNQNDKVINVNGAAHKNQ 767
QY 743 AYRPLILGTDGVQSVLYKSDST-NIVKYTDANGNLTFADDIKGYSTVDSMGVAVWPV 801

DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:| 768 EYRPLLITTKOGLTSYTSDAAKSLYRXTNDKGLVFDASDIQGYLNPQVSVGLAVWPV 827
QY 802 GAKGQDQYRVAADTNQKADGKSLKTSAAALDSQVIYEGFSNFQDFANNADATYTNKKAENA 861
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:| 828 GASDNQDQYRVAASNKANATQGVYESSALDSQLIYEGFSNFQDFVTKDSYTNKKAQNV 887
QY 862 DFFKLGITSEMAPQYVSATDGSFLDSIIQNGYAFSDRYDLAMSKNNKYSKDDLANAL 921
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:| 888 QLFKSWGVTSEMAPQYVSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYSQQQDMINAV 947
QY 922 KALHANGIQAATADWVPDQIYOLPGBEVVTAKRTNSYGNPTEDAYINNALYANTNKSQSD 981
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:| 948 KALHKSGLQVIAWVPDQIYNLPKBEVVTATRVNDYGEYRKDSEIKNTLYAANTKSGKD 1007
QY 982 YQAYGGAFDELKAKYDPMFTVMNMISTGKPIDPSTKI KQWEAKYFNGTNVLGKAGYVL 1041
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:| 1008 YQAYGGAFLSLAAYKPSIFNRTQISNGKIDPSEKITAWKAKYFNGTNILGRGVYVL 1067
QY 1042 SDDATGKYFTVNBNGDFLPASFTGDONAKTGYDYDGTGMAYYSTSGNKAIVNSFIYEG-GH 1100
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:| 1068 KDNASDKYFELKGNQTYLPKQWT-NKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKN 1126
QY 1101 YYPFDKQCHMTGSKAEDGNDYVFLPNGIOWRDAIYODAGNSYYYGRTGLYKGDNMY 1160
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:| 1127 WYFDNNGHMVGLOQL-NGEVOIFLSNGVQURSFLENADGSKNYFGLGNRY-SNGYY 1184
QY 1161 PFVDNNANKTVFRYFDANNYMAIGYRNMYGQTYFYFDENGFGQAKQLLTDKKG-THYFDE 1219
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:| 1185 SF-----DNDSKWRVFDASGVMAVGLKTINGNTQYFDQDGYQVKGAWITGSDGKKRYFDD 1239
QY 1220 DNGAMAKKFN-VGDDWYMDGNGNAVKQYVNNQILYFNPEYGVQVKGQFIIDAQR 1278
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:| 1240 GSGNMAVNRFANDKNGDWYILNSDGIALVGQTINGKTYFYFG-QDGKQIKGKIITD-NGK 1297
QY 1279 TSYVDANSALKSGGFETPNGSDWYYAENGYYVYKQVAENQDQWYFDDTGTQKQKA 1338
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:| 1298 LKYFLANGELARNIFAT-----DSQNNWYIFG-SDGVAVTGS 1334
QY 1339 AKVGDRLYFNPDSGVQVKGDFATDESNTSPYHGDNGDKVVGGFPTTGNNAWYYADNNG 1398
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:| 1335 QTIAGKKLYFASD-GKQVKGFSVT-YNGKVHYHADSGELQVNRFEADKDGNNWYLDNSG 1392
QY 1399 NLVKGFBQIDGKWHYHDFDEVTCQQAAGALVNGQQLYFVDVDSGIQVKGDFVTD 1450
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:| 1393 EALTGSORI-----NDQRVFF-TREGKQVKGDVAYD 1422
RESULT 9
AAU98042
ID AAU98042 standard; protein; 1430 AA.
XX
AC AAU98042;
XX
DT 27-AUG-2002 (first entry)
XX
DE S. mutans glucosyltransferase GFPD mutant T589E.
XX
KW Glucosyltransferase; GFPD; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture; mutant; mutin.
OS Streptococcus mutans.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 589 /note= "Wild-type Thr substituted by Glu"
XX
PN US2002031826-A1.
XX
PD 14-MAR-2002.
XX

QY 462 ADLLQVSYVFNKYNKYVTDSEANALAHISILEAWSLNDNQYNEBTDNGTALSIDNSRLTFS 521
 Db 472 ADLLQIASDYLYKAHYGVKSEKNAIINHLSLEAWSNDPQYNKDTKGAQLPDKLRLSL 531
 QY 522 LAVLTQPCORIDLSN-----LISESVNKERANDTAYGDTTPTYSFVRAHSEVQ 571
 Db 532 LVALTR-PLEK-DASWNEIRSGLEPVTITNSLM-NRSABGKNSERWANYIFIRAHSEVQ 588
 QY 572 TVIAKIVKEKIDTNSDGYFTTLDQDAFKIYNEDMAKNKTYTHYNIIPAAYALLISNME 631
 Db 589 DVIAKIIKAQINPKTDGLTGLTDELQAQAFKIYNEDMRQAQKYYTQSNITAYALMLSKD 648
 QY 632 SVPRVYGYLDYDDGQYMAKSPYDAIATMLGRIATYVSGGSEEVHKVNGNQ----- 686
 Db 649 SITRLYGYMYSDDGYMATKSPYDAIDITLLKARIKAAAGQDMKITTYVEGDKSHMDW 708
 QY 687 ---ILSSVRYGDLMSADITQGTDLRSGLTVLVSNDPDLG-GDSLTVNGRAHANQ 742
 Db 709 YTGVLTSVRYGTGANEATD-QGSEATKTQGMVITSNPFLKLNQNDKVINVMGAHKNQ 767
 QY 743 AYRPJLTGKDGVSQSYLKDSDT-NIVKYTDANGNLTFTADDIKGYSTVDMSGYLAWVVPV 801
 Db 768 EYRPLLTTKDLGTSYDAASAKSLYKTDNGELVFDASDIQGYLNPQVSGYLAWVVPV 827
 QY 802 GAKGDQVRVAADTNQKADGKSLKTSALDSQVITYEGFSNFQDFANNDADYTNKKIAENA 861
 Db 828 GASDNQDVRVAASNAKANATQGVYESSALDSQLIYEGFSNFQDFVTKDSYTNKKIAQV 887
 QY 862 DFEKGLGTSPEMAPOYVSATGSGFLDSIIQNGYAFSDRYDLAMSNNKYGSKDDLNAL 921
 Db 888 QLFKSGWTSFEMAPQYVSSEDSGLDSIIQNGYAFSDRYDLAMSNNKYGSKQDMINAV 947
 QY 922 KALHANGIOAIADWPQDQIYQLPGEVVTAKTNSYGNPTFDAYINNALYANTKSSGSD 981
 Db 948 KALHSGIOVIADWPQDQIYQLPGEVVTATRVNDYGEYRKDSEIKNTLYAANTKNSGKD 1007
 QY 982 YQAYGGAFDLBELKAKYPMFTVMNISTGKPIDPSTKIKQWBAKYPNGTNVLGKAGYYL 1041
 Db 1008 YQAYGGAFDLSELAAYPSIFNRTQISNGKIDPSEKITAFAKAYFNGTNILGRGVYV 1067
 QY 1042 SDDATGKYFTVNEGDFLPASTGTQONAKTYGYGTGMAYYSTSGNAKAVSFYEG-GH 1100
 Db 1068 KDNASDKYFELKGNQTYLPKQMT-NKEASTGFVNDGNGMTFTYSTGYQAKNSFVQDAKN 1126
 QY 1101 YYPFDKGHMTGVSYKAEDGNDYVFLPNCIQMRDAIYODACGNSVYVYGRGTLYKGDNMY 1160
 Db 1127 WYFPDNGHMYVGLQOL-NGEVOYFLSNGVQVRESFLENADGSKNYFHLGNRY-SNGY 1184
 QY 1161 PFVDPNNANKTVRYFDANNVMAIGYRNMYGQTYFYDENGFOAKQLLTDKRG-THYFDE 1219
 Db 1185 SF-----DNDSKWRVYFDASGVMAVGLKTINGNTQYFDQDGYQVKGAWITSGDKKRYFDD 1239
 QY 1220 DNGAMAKKFNV-VGDDWYMDGNGNAVKGPVANNQILYFNPTGVQVKGQFITTDAQR 1278
 Db 1240 GSGNMAVRFANDKNGDWYLLNSDGIALVGQTINGKTYFYG-QDGKQIKGKIITD-NGK 1297
 QY 1279 TSYDYDANSALKSSGFFTPNGSDWYVYKGFQVAENQDQWYVFDQTTQKQAKA 1338
 Db 1298 LKYLFIANGSELARNIFAT-----DSQNNWYFYG-SDGVAVTGS 1334
 QY 1339 AKVGDRLYFNPDSGVQVKGDFATDESIGNTSFYHGDNGDKVYGVGFFTTGNNAWYADNNG 1398
 Db 1335 QTIAGKLYFASD-CKQYKGSFVT-YNGKVHYHADSGELQVNRFEADKDGWYLLDSNG 1392
 QY 1399 NLVKGQFQIDGKWHYHDFEVTGQQAAGALVNGQQLYFVDVDSIGYVKGQFVTD 1450
 Db 1393 EALTGSQRI-----NDQRVFF-TREGQVKQGDVAYD 1422

RESULT 11
 AAU98044
 ID AAU98044 standard; protein; 1430 AA.
 XX

AC AAU98044;
 XX 27-AUG-2002 (first entry)
 DT
 XX S. mutans glucosyltransferase GTFD mutant N471D/T589D.
 DE
 XX Glucosyltransferase; GTFD; transgenic plant; paper sizing;
 KW coating composition; glucan; starch; latex; thermoplastic molecule;
 KW amyloplast; vacuole; paper manufacture; mutant; mutein.
 XX
 OS Streptococcus mutans.
 OS Synthetic.
 XX
 XX Key Location/Qualifiers
 FH Misc-difference 471 /note= "Wild-type Asn substituted by Asp"
 FT Misc-difference 589 /note= "Wild-type Thr substituted by Asp"
 FT
 XX US2002031826-A1.
 PN
 XX 14-MAR-2002.
 PD
 XX 19-DEC-2000; 2000US-00740274.
 PF
 XX 07-JUN-1995; 95US-00478704.
 PR 07-JUN-1995; 95US-00482711.
 PR 07-JUN-1995; 95US-00485243.
 PR 16-JAN-1998; 98US-00007999.
 PR 16-JAN-1998; 98US-00008172.
 PR 20-JAN-1998; 98US-00009620.
 PR 11-DEC-1998; 98US-00210361.
 XX (NICH/) NICHOLS S E.
 PA
 XX Nichols SE;
 PI
 XX WPI; 2002-414332/44.
 DR
 XX
 PT Glucosyltransferase B or D protein useful for producing a glucan useful
 as substitutes for and additions to modified starch and latexes in paper
 PT manufacture, comprises mutations in specific positions.
 XX
 PS Claim 36; Page; 44pp; English.
 CC
 XX The invention an isolated protein comprising a glucosyltransferase (GTF)
 CC B polypeptide having changes at position from 1448V, D457N, D567T,
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/D571K/K1014T,
 CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
 CC GTF mutant, an isolated polynucleotide which encodes PI or P2, or its
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the
 CC GTF mutant, an expression cassette comprising the polynucleotide operably
 CC linked to a promoter, a vector comprising the expression cassette, host
 CC cell introduced with the vector, a transgenic plant comprising the
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
 CC coating composition comprising a glucan produced in a plant transformed
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
 CC thermoplastic molecule or their combinations or glucan and starch where
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line
 CC deficient in starch biosynthesis, transformed with a gene encoding a
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
 CC comprising the glucan (paper sizing/coating agent). The vector is useful
 CC for producing a glucan in a plant. The method comprises transforming a
 CC plant cell with the vector, growing the plant cell under plant growing
 CC conditions to produce a regenerated plant and inducing expression of the
 CC polynucleotide for a time sufficient to produce the glucan in the
 CC regenerated plant, where the vector contains a transit sequence from
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.
 CC Glucans are useful as substitutes for and additions to modified starch

CC and latexes in paper manufacture. Unlike prior art techniques, which
CC require input materials that produce chemical effluents, paper
CC manufacture utilising the glucan produced by GFP, which utilises
CC biologically produced input materials, is more cost-effective and
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
CC properties and impart gloss to the paper during coating step. The present
CC sequence represents a GFPD mutant of the invention. Note: The present
CC sequence is not shown in the specification and was created by the indexer
CC using the GFPD sequence appearing as AAU98029 and the information in
CC claim 36
XX
SQ Sequence 1430 AA;

Query Match 43.8%; Score 3604; DB 5; Length 1430;
Best Local Similarity 49.8%; Pred. No. 3.4e-182;
Matches 743; Conservative 218; Mismatches 419; Indels 112; Gaps 32;
QY 1 MEKLLHYKLHKVKKHWTIAVASIGLVSLVGAGTVSAEDKVDANDTTAQTAVGVDTCQ--- 57
DB 1 METKRYKMHKVKHWTIVAVAS-GLITL---GTTLGSSVSAETEQTQSDKVVTQKSD 56
QY 58 DQATTNDANTNTDITDQASANTNQDQAGSQNNQDQAKQDT-----ANTDR 106
DB 57 DKAASESSQTDAPKT---KQATQEQTAQS-QANVADTSTITKETPSQNTTQANSDD 111
QY 107 NOADNSQTDNNQATQOATSPTDGTSGVORRDAANVATA-ADQEGOTAPSEQKSAAL-SL 164
DB 112 KTVNTNKSEBAEQTBERTKQSEBAQTASSQALTOAKAELTKQRTAAQENKNPVDLAAI 171
QY 165 DNVKLIDGKYYVQADGSKYKFNALTVNGOMLYFSDTGALSTSTSYFSQGTNLVDPF 224
DB 172 PNVKQIDGKYYIGSDGQPKNFALTVNNKVLVFDKNTGALTDTTQYQPKQGLTKLNNDY 231
QY 225 SSHNKAYDSTAKSFELVNGYLTANSWYRPAGILRNGOTWEASNEENDLRPLMSWPPDKOT 284
DB 232 TPNQIVNFPENTISLETIDNYVTADSWYRPKDLKNGKWTWASESDLRPLMSWPPDKOT 291
QY 285 QVAYVNMNKLILSANETEVNTSQTVDLNEAQSITQKIEQKITSDNSTQWILRTAMEAPV 344
DB 292 QIAYLNMNQOGLGTGENYTADSSQESLNLAAQTQVKIETKISQTQOTQWLDRDIINSFV 351
QY 345 AAQPKWNWSTE---NFNKGDLHOGGALLYTNSDLTPWANSDYRLNRTPTQODGTKKYFT 401
DB 352 KTOPNWSQTESDTSAGEKDHIOGGALLYNSDKTAYANSDYRLNRTPTSTQTKPKPYE 411
QY 402 EGEGGYEFLNDVNSNPVQAEOQLNQLHLYLMNWGDIVMGDXDANFDGVRVDAVDN 461
DB 412 DNSSGGYDFLLANDIDNSNPVQAEOQLNQLHLYLMYGSIVANDPEANFDGVRVDAVDND 471
QY 462 ADLLQVSNYFKDNYKVTDSEANALAHSTILSAWSLNDNQYNEDTNGTALSIDNSRLTS 521
DB 472 ADLLQIATADYLKAHYGVDSKXNAINHLSLEAWSNDNPQYNKDTKGAQLPIDNKLRLSL 531
QY 522 LAVLTKQPCQORDLSN-----LISEVNKBERANDTAYGDTIPTYSFVRADHSEVQ 571
DB 532 LVALTRE-PUK-DASNNKNIIRSGLEPVITNSLN-NRSEAGKNSERMANVIFIRAHDSVQ 588
QY 572 TVIAKIVKEKIDTNSDGYFTFLDQKDAFKIYNEDMAKVNKYTHYNIIPAAVALILSNME 631
DB 589 DVIAKIIKAIQINPKTDGLTFLDELKQAFKIYNEDMRQAKKYTQSNITPATALMLSNKD 648
QY 632 SVPRVYGLYTDGQYMAKSPYYDAIATMLQGRYAYVSGQSEVHKVNGNQ----- 686
DB 649 SITRLYVGMYSDDGQYMATKSPYYDAIDTLKARIKYAAGQDMKITVYEGDKSHMDWD 708
QY 687 ---ILSSVRYGQDLMSADTQGTDLRSRTSLVTLVSNDPNLDL-GDSLTVNMGRAHANO 742
DB 709 YTGVLTSVRYGTGANEATD-QGSEATKTQGMVITSNPSSLKLNQNDKVINVMGAHRNQ 767
QY 743 AYRPLILGTGQGVQSYLKSDSDT-NIVKYYTDAMGNLTFTADDIKGYSTVDMSGYLAVWVPV 801
DB 768 EYRPLLLTTKDLGTSYTSDAAKSLYRKNTDKGELVFDASDIQGYLNPQVSGYLAVWVPV 827

QY 802 GAKQGDVRYAADTNQKADGKSLKTSAAALDSQVIYEGFSNFODFANNADYTNKKIAENA 861
DB 828 GASDNQDVRVAASNKANATGOVYESSALDSQLIYEGFSNFODFVTKDSYTNKKIAQNV 887
QY 862 DFFKKLIGTSPEMAPQYVSATDGSFLDSIIQNGYAFSDRYDLAMSKNNKYGSKDDLALAN 921
DB 888 QLFPKSWGVTSEMAPQYVSSEDSFLDSIIQNGYAFEDRYDLAMSKNNKYGSKQODMINAV 947
QY 922 KALHANGIOAIADWVPDQIYOLPGEVVTAKRTNSYGNPTFDAYINNALLYANTWSSGSD 981
DB 948 KALHKSIGIQTADWVPDQIYNLPGEVVTATRVNDYGEYRKDSEIKNTLYAAANTKSNKGD 1007
QY 982 YQAOYGGAFDELKAKYDPMFTVMNISTGKPTDPTKIQWEAKYFNGTNVLGKAGVYL 1041
DB 1008 YQAYGGAGFLBELAAKYPISIFNRTQISNGKIDPSEKITAWKAKYFNGTNILGRGVGYL 1067
QY 1042 SDDATGKYFTVNEGDFLPASFTGQNAKTGFYDGTGMAYYSTSGNKAVNSFIYEG-GH 1100
DB 1068 KDNASDKYFELKGNQTYLPKQWT-NKEASTGFVNDGNGMTFYTSTGYQAKNSFVQDAKN 1126
QY 1101 YYYFDKGHWMTGSKYABEDGNDYFELPNGIQMRDAIYQDAQNSYYYGRTGILYKGNWY 1160
DB 1127 WTYFDNNGHVVYGLQOL-NGEVQYFSLNGVQLRESFLENADGSKNYPFGLGNRY-SNGYY 1184
QY 1161 PFVDNNANKTVFRYFDANNVMAIGYRNMYGOTYYFDENGFQAKGQLLTDGK-THYPDE 1219
DB 1185 SF-----DNDSKWRYFDASGVMAVGLKTINGNTQYFDQDGYQVKGAWITGSDGKKRYFPD 1239
QY 1220 DNGAMAKNKFVN-VGDDWYMDGNGNAVKGQYFVNQILYFNPETGTVQVKGQFPTDAQGR 1278
DB 1240 GSGNMAVNRFPANDKNDWYLLNSDGIALVGQTINGKYYFG-QDGKQIKGLIITD-NGK 1297
QY 1279 TSYDANSALKSSGFFTPNGSDWYYAENGVYVYKGFQVAENQDQWYYPDQTTGQAKGA 1338
DB 1298 LKYFLANGSGLARNIFAT-----DSQNNWYFG-SDGVAVTGS 1334
QY 1339 AKYDGRDLYFNPDSGVQVKGDPATDESNTSFYHGDNGDKVVGFFTTGNNAWYADNNG 1398
DB 1335 QTIAGKLLYFASD-GKQYKGSFVT-YNGKVHYHADSGELQVNRFEADKGNWYLDNSG 1392
QY 1399 NLVKFQFQIDGKWHYFDEVTGQAKGAALVNGQQLYFVDVDSGIQVKGDPVTD 1450
DB 1393 EALTSQRI-----NDQRVFP-TREGKQVKGDVAYD 1422
RESULT 12
AAU98045
ID AAU98045 standard; protein; 1430 AA.
XX
AC AAU98045;
XX
DT 27-AUG-2002 (first entry)
XX
DE S. mutans glucosyltransferase GFPD mutant N471D/T589E.
XX
KW Glucosyltransferase; GFPD; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture; mutant; mutein.
OS Streptococcus mutans.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 471
FT /note= "Wild-type Asn substituted by Asp"
FT Misc-difference 589
FT /note= "Wild-type Thr substituted by Glu"
XX
PN US2002031826-A1.
XX
PD 14-MAR-2002.
XX
PF 19-DEC-2000; 2000US-00740274.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 19:39:30 ; Search time 34.0237 Seconds
(without alignments)
3776.130 Million cell updates/sec

Title: US-10-797-821-38

Perfect score: 8237

Sequence: 1 MEKLUHYLHKYKHVWTTIA.....FVRIGDQWYFGNDGAATNL 1554

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/prodata/1/iaa/5 COMB.pep.*
2: /cgn2_6/prodata/1/iaa/6 COMB.pep.*
3: /cgn2_6/prodata/1/iaa/H COMB.pep.*
4: /cgn2_6/prodata/1/iaa/PCTUS COMB.pep.*
5: /cgn2_6/prodata/1/iaa/RE COMB.pep.*
6: /cgn2_6/prodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3615	43.9	1430	2	US-09-008-172-2
2	3615	43.9	1430	2	US-09-210-361-6
3	3615	43.9	1430	2	US-09-740-274-6
4	3434.5	41.7	1475	2	US-09-007-999-2
5	3434.5	41.7	1475	2	US-09-210-361-2
6	3434.5	41.7	1475	2	US-09-740-274-2
7	3429.5	41.6	1577	1	US-08-793-824-2
8	3161.5	38.4	1375	2	US-09-210-361-4
9	3161.5	38.4	1375	2	US-09-740-274-4
10	2422	29.4	2057	2	US-09-499-203-2
11	2285	27.7	1781	2	US-09-604-957-3
12	2285	27.7	1781	2	US-09-995-749A-2
13	1643	19.9	545	2	US-09-604-957-4
14	1615	19.6	545	2	US-09-995-749A-10
15	1442.5	17.5	522	2	US-09-995-749A-11
16	1440	17.5	523	2	US-09-604-957-5
17	1365	15.4	535	2	US-09-604-957-7
18	1365	15.4	535	2	US-09-995-749A-13
19	1174.5	14.3	584	2	US-09-604-957-6
20	1172.5	14.2	584	2	US-09-995-749A-12
21	563	6.8	349	2	US-09-009-620-2
22	545.5	6.6	2710	1	US-08-480-604A-6
23	545.5	6.6	2710	1	US-08-405-496A-6
24	545.5	6.6	2710	2	US-08-915-136-6
25	545.5	6.6	2710	2	US-08-957-310-6
26	545.5	6.6	2710	2	US-08-011-366-6
27	545.5	6.6	2710	2	US-09-084-517-6

28	469	5.7	866	2	US-09-545-773-2	Sequence 2, Appli
29	469	5.7	866	2	US-10-222-038-2	Sequence 2, Appli
30	465	5.6	1231	2	US-08-714-741-41	Sequence 41, Appl
31	463	5.6	2366	1	US-08-480-604A-10	Sequence 10, Appl
32	463	5.6	2366	1	US-08-405-496A-10	Sequence 10, Appl
33	463	5.6	2366	2	US-08-915-136-10	Sequence 10, Appl
34	463	5.6	2366	2	US-08-957-310-10	Sequence 10, Appl
35	463	5.6	2366	2	US-10-011-366-10	Sequence 10, Appl
36	463	5.6	2366	2	US-09-084-517-10	Sequence 10, Appl
37	458	5.6	811	1	US-08-480-604A-7	Sequence 7, Appli
38	458	5.6	811	1	US-08-405-496A-7	Sequence 7, Appli
39	458	5.6	811	2	US-08-915-136-7	Sequence 7, Appli
40	458	5.6	811	2	US-08-957-310-7	Sequence 7, Appli
41	458	5.6	811	2	US-10-011-366-7	Sequence 7, Appli
42	458	5.6	811	2	US-09-084-517-7	Sequence 7, Appli
43	458	5.6	812	1	US-08-480-604A-29	Sequence 29, Appl
44	458	5.6	812	2	US-08-915-136-29	Sequence 29, Appl
45	458	5.6	812	2	US-09-084-517-29	Sequence 29, Appl

ALIGNMENTS

RESULT 1

US-09-008-172-2
; Sequence 2, Application US/09008172
; Patent No. 6127602
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0358D
; CURRENT APPLICATION NUMBER: US/09/008,172
; CURRENT FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-008-172-2

Query Match	43.9%	Score	3615	DB 2	Length	1430
Best Local Similarity	49.9%	Pred. No.	9.7e+211			
Matches	745	Conservative	217	Mismatches	418	Indels 112; Gaps 32
QY	1	MEKLUHYLHKYKHVWTTIAVASIGLVSLVGAGTVSAEDKVANDTTAQATVGVDTCQ---	57			
Db	1	METKERYKMHKVKHVVAVAS-GLITL---GTTTGGSSVSAETEQQTSKVVTKQSD	56			
QY	58	DOATNDANTNTTDTTADQASANTNQQDQSGSQSNQDQAKQDT-----ANTOR	106			
Db	57	DKAASESSQTDAPKT---KQAQTEQTQAQS-QANVADTSTTITKTPSQNTTQANSDD	111			
QY	107	NOADNSQTDNNQATQATSPADGTGTVQRDAANVATA-ADQEGOTAPSEQEKSAAL-SL	164			
Db	112	KTVNTKSEEAQTSSERTKQSEEAQTASSQALTOAKAELTKQRTAAQENKNFVDLAAI	171			
QY	165	DNVKLIDGKYVYVQADGSKKNFALTAVAGQMLYFSDTGALSTSTYSFSQGTNLVDDF	224			
Db	172	PNVKIDGKYVYIGSDGQPKKNFALTAVNKKVLYFDKNTGALTDTSQYQPKQGLTLNNDY	231			
QY	225	SHNKAYDSTAKSFELVNGYLTANSWYRPAQILRNQGTWEASNEENDLRPLVLSWPPDKDT	284			
Db	232	TPHNQIVNFENSTLETDINYYVTADSWYRPAQILRNQGTWEASNEENDLRPLVLSWPPDKDT	291			
QY	295	OVAVYVYNNKYLSANETETVNTSETSOVDNKEAQSTQTKLEQKITSNDSNTOWLRTAMEFV	344			
Db	292	QIAYLYNNQOGLGTGENYNTSSQESLNLAQQTVQVKIETKISQQTQOWLRDIINSFV	351			
QY	345	AAQPKWNMSTE---NFNKGDLHGGALLYTNSDLTPWANSYRLLNRTPTQDGTKKYFT	401			

```
Db      352  KTQPNWNSQTESDTSAGEKDHLOGGALLYNSDKTAYANSDFRLNRLNPTSTQGPKE 411
Qy      402  EGGEGYEFLLSNDVNSNPVQAQBNOLHYLMNMWDIVMGDKDANFDGVRVADVNN 461
Db      412  DNSSGGYDFLLANDIDNSNPVQAQBNOLHYLMNYSIVANDPEANFDGVRVADVNN 471
Qy      462  ADLLQVSNFYKDYKVTDSANALAHISILEAWSLNDQYNEDTNGTALSIDNSRLTS 521
Db      472  ADLLQIASDYLKAHYGVDSKSEKNAHLSILEAWSNDPQYNKDTKGALPLIDNKLRLSL 531
Qy      522  LAVLTMPORIDLSN-----LISEVNERANDTAYDGTIPTVSFVRADSEVQ 571
Db      532  LVALT-R-PLEK-DASKNKIRSLGEPVITNSLN-NRSAGKNSRMANYIFIRADSEVQ 588
Qy      572  TVIAKIVREKIDNSDGYFTFLDQLDAFKIYNEDMAKYNKTYTHYNIPAAVALLSNME 631
Db      589  TVIAKIIKAQINPKTDGLTFLDELKQAFKIYNEDMRQAKKYTQSNIPTAYALMSNKD 648
Qy      632  SVPRVYVYGDLYTDDQYMAKKSPPYDAIATMLQRIAYVSGQSEVHKVNGNQ----- 686
Db      649  SITRLYVYGDYVSDGQYATKSPYYDAIDITLLKARIKYAAGQDMKITVVEGDKSHMDWD 708
Qy      687  ---ILSSVRYGDLMSADDTQGTDLRSRTSLVTLVNSDNLDLG-GDSLTVNMGRAHNQ 742
Db      709  YTGVLTSVRYGTGANEATD-QGSEATKQGMVITSNPFLKLNQNDKVINVMGAAHRNQ 767
Qy      743  AYRPLILTKDGVQSYLKDSDT-NIVKYTDANGNLFTTADDIKGYSTVDMSGYLAVWVPV 801
Db      768  EYRPLLLTTKDLGTSYTSDAAKSLYRKTDKGELVDFDASDIQGYLNPQVSGYLAVWVPV 827
Qy      802  GAKGDQVRVAADTQKADGKSLKTSALDSQVITYEGFSNFQDPANNADYTNKKIAENA 861
Db      828  GASDNQDVRVAASNKANATQGVYESSALDSLIYEGFSNFQDPVTKDSDYTNKKIAQNV 887
Qy      862  DFFPKLGITSPMAQVYVATDGSFLDSTIQNGYAFSDRYDLAMSKNNKYSGKDLANAL 921
Db      888  QLFKSGVTSFEMAQVYVSESDGSFLDSTIQNGYAFPEDRYDLAMSKNNKYSGQDMINAV 947
Qy      922  KALHANGIOAIADWVPDQIYQLPGEVVTAKRTNSYGNPTFDAYINNALYATNTKSSGSD 981
Db      948  KALHKSQIQIADWVPDQIYNLPGEVVTATRVNDYGEYRKDSEIKNTLYAANTKSKNGD 1007
Qy      982  YQAQYGGAPLDELKAKYDPMFTVNMISTGKPIDPSTKIKQWEAKYFNGNINVLGKAGYVL 1041
Db      1008  YQAQYGGAPLSELAARYPSIFNRTQISNGKIDPSEKITAWKAKYFNGNINILGRGVYVL 1067
Qy      1042  SDAATGKYFTVNEGDFLPASFTGQNAKTGFYDGTGMAYYSTSGNKAVNSFIYEG-CH 1100
Db      1068  KDNASKYFELKGNQTYLPKQMT-NKEASTGFVNDGNGMTFTSTSGYQAKNSFVQDAKN 1126
Qy      1101  YYYFDKDHMTVSGYKAEDGNDYFLPNGIQMRDAIYDQAQNSYYYRGRTGILYKGNWY 1160
Db      1127  WYFDNNGHMVYGLQL-NGEVQYFLSNGVQLRESFLENADGSKNYFGLGNRY-SNGY 1184
Qy      1161  PFVDPNNANKYFTRFDANNVMAIGYRNMVGYTFYFDENGFOAKCOLLTDGK-THYFDE 1219
Db      1185  SF-----DNDSKRWYFDASGVMAVGLKTINGNTQYFDQGYQVKGAWITGSDGKRYFDD 1239
Qy      1220  DNGAMAKNKFVN-VGDDWYTMDBGNGNAVKGOYPVNNOILYFNPETGVQVKGOFITDAQCR 1278
Db      1240  GSGNMAVNRFPANDKNGDWYLLNSDGLALVGQVINGKTYFG-QDGKQIKGLIITD-NGK 1297
Qy      1279  TSYDANSALKSSGFFTPNGSDWYIYAENGYYVYKGPQVAENQDQWYFDTQTGQAKGA 1338
Db      1298  LKYFLANSSELARNIFAT-----DSQNNWYFYG-SDGVAVTGS 1334
Qy      1339  AKVDGRDLYFNPDSGVQVKGDFATDESGMTSPFHGDNQDKVYGGFFTTGNNAWYADNNG 1398
Db      1335  QTIAGKKLYFASD-GKQVKGVSFT-YNGKVHYHYHADSGELQVNRFPADKDGWYLLDSNG 1392
Qy      1399  NLVKGFQEIADGKWYHDFEVTGQQAAGALVNGQLYFVDVDSIGIVKQGDFTVD 1450
```

```
Db      1393  EALTGSQRI-----NDQRVFF-TREGKQVKGDVAYD 1422

RESULT 2
US-09-210-361-6
; Sequence 6, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-210-361-6

Query Match      43.9%; Score 3615; DB 2; Length 1430;
Best Local Similarity 49.9%; Pred. No. 9,7e-211;
Matches 745; Conservative 217; Mismatches 418; Indels 112; Gaps 32;

Qy      1  MEKKLHYKLHKYKHWVTIAVASIGLVSIVGAGTVSABDKVANDTTAQTAVGVDGQ---- 57
Db      1  METKRRYKMHVKCHWTVAVAS-GLITL--GTTTLGSSVSAETEQTSDKVVTKQSD 56
Qy      58  DOATINDANTTDTADQSANNTNODAGSDOSNNQDAOKDT-----ANTDR 106
Db      57  DKAASESQTDAKPT-----KQAEQTQQAQS-QAANVDTSTSIKETPSQNTITQANSDD 111
Qy      107  NOADNSQDNNQAQDQATSPATDGTSVQRRDAANVATA-ADQEGQTAPSEQEKSAAL-SL 164
Db      112  KTVTNTKSEEAQTSERTKQSEEAQTASSQALTQAKAELTKQRTAAQENKNPVDLAAI 171
Qy      165  DNVKLIDGKYIVQADGSKYKKNFALTIVNGQMLYFSDTGALSSSTYSFSQGTTLVDDF 224
Db      172  PNVKQIDGKYIYIGSDGQPKKNFALTIVNNKVLVFDKNTGALTDTTSQYQPKQSLTKLNNDY 231
Qy      225  SSHNKAYSTAKSFEELVNGYLTANSWYRPAGLIRNGQTWASNSNENDLRPLVMSWPKDT 284
Db      232  TPNQIVNFENTSLETIDNYVTADSWYRPKDILKNGKWTWASSEDRLRPLMSWPKDT 291
Qy      285  QVAYVYNNKVLNSANETEVNTNETSQVDLNLKQAQSIQTKEQKITSNDSNTQWLRTAMEAFV 344
Db      292  QIAYLNYNNQGLGNGENTADSSQESLNLAQTQVVKIETKISQTQQTQWLRIINSFV 351
Qy      345  AAQPKWNMSTE---NFNKGDLHQQGALLYTNSDLTPWANSYRLNRLNRTPTQDGTGKYTF 401
Db      352  KTQPNWNSQTESDTSAGEKDHLOGGALLYNSDKTAYANSDFRLNRLNRTPTSTQGPKE 411
Qy      402  EGGEGYEFLLSNDVNSNPVQAQBNOLHYLMNMWDIVMGDKDANFDGVRVADVNN 461
Db      412  DNSSGGYDFLLANDIDNSNPVQAQBNOLHYLMNYSIVANDPEANFDGVRVADVNN 471
Qy      462  ADLLQVSNFYKDYKVTDSANALAHISILEAWSLNDQYNEDTNGTALSIDNSRLTS 521
Db      472  ADLLQIASDYLKAHYGVDSKSEKNAHLSILEAWSNDPQYNKDTKGALPLIDNKLRLSL 531
```

QY	522	LAVLTQPCQORIDLSN-----LISESVNKERANDTAYGDTIPYTSFVRAHDSVQ	571
Db	532	LVALTR-PLEK-DASNKNEIRSGLEPVTJNSLN-NRSAEGKNSESMANYIIFRAHDSVQ	588
QY	572	TVIAIKVKEIDTNSDGYTFTLDQLKDAFKIYNEDMAKVNTYTHYNI:PAAYALLLSNWE	631
Db	599	TVIAIKIQAQINPKTDGLTFTLDELKQAFKIYNEDMRAKKYQTSQNI:PTAYALMLSNKD	648
QY	632	SVPRVYVYGDLYTDDGOYMAKSPYYDAJATMLQGR:IAYSVGGQSEBHVKAVNGNQ-----	686
Db	649	SITRLYYGDMYSDDGOYMATKSPYYDA:IDTLLKARIKYAAGQDMKITYVEYGDKSHMDWD	708
QY	687	---ILGSVRYGQDLSMSADDTQGTDLSPRTSGELVTLVSNDPNLDLG-GDSLTVMGGRAHANQ	742
Db	709	YTGVLTSVRYGTGANEATD-QGSEATKTCQMAVITSNP:SLKLNQNDKVIYNMGAHKNQ	767
QY	743	AYRPLILGTQDQVQSYLKDSDT-NIVKYT:TDANGNLTF:TADDIKGYSTVDMSGYLAVMVPV	801
Db	768	EYRPLLLTTKOGLTSYTSDDAAKSYIRKTNKDELVF:DA:SDIQGLPNQVSGYLAVMVPV	827
QY	802	GAQGDQDVRVAADTNQKADGSKLKTSAALDSQV:YEGFSNFQDFANNDADYTNKKIAENA	861
Db	828	GASDNQDVRVAASNKANATGOVYESSALDSQL:YEGFSNFQDFVTKDSDYTNKKIAQNV	887
QY	862	DFFKLGTITSEMAPQYVSATDGSFLDSIIQNGYAFSDRYDLAMSKNKYSGKDDLANAL	921
Db	888	QLFKSWGVTSEMAPQYVSSDGSFLDSIIQNGYAFEDRYDLAMSKNKYSGSQDMINAV	947
QY	922	KALHANGIOA:ADWVPDOIYQ:PEGEVVYAKRTNSYNGNPTPDAYINNALYATNTKSSGSD	981
Db	948	KALHKSGIQV:ADWVPDOIYNLPKGEVVYATRVNDYGEYRKDSB:IKNTLYAANTKSNXKD	1007
QY	982	YOAQYGGAFLELDELKAKYPMFTVNMIS:TGKPIDPSTKIKOWEAKYENGTVNLHGKAGVVL	1041
Db	1008	YQAKYGGAFLELAKYPSI:FNRTQISNGKKIDPSEKITAWAKYFNGTNILHGRGVGYL	1067
QY	1042	SDDATGYFTYVNGNDELFPASFTGDQNAKTFYFDGTGMAYYSTSGNKAVNSFYEG-GH	1100
Db	1068	KDNASDKYFELKGNQYLYPKQMT-NKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGN	1126
QY	1101	YYYFDPKQGHMVTGYSKAEADGNDYFLPNGIQMRDAIYQDAQNSYYYGRTGILYKGDNWY	1160
Db	1127	WYYFDNNGHMYGLQQL-NGEVQYFLSNGVQLRESFLENADGSKNYFEGHLGNRY-SNGYY	1184
QY	1161	PFVDPNNAKTVFYRYFDANNVMAJLYRNMWYQTYFYFBNQFPAQQLLTDQKG-THYFDE	1219
Db	1185	SF-----DNDSKMWRYFDASGMVAGLKTINGNTQYFDQDGYQVYKGAWITGSDGKKRYFDD	1239
QY	1220	DNGAMAKNKFVN-VGDDWYVMDGNGNAVKGOYPVNNQILLYFNPETGVQVKGQFTTDAQGR	1278
Db	1240	GSQNMVNRFANDKNGDWYVYLLNSOGIALUVQVTTINGTKTYFYG-QDQKQIKGKIITD-NGK	1297
QY	1279	TSYYDANSALKSSGFFTPNGSDWYVYKGFQKQVAENQDWYFYFQTTGKQAKGA	1338
Db	1298	LKYFLANSGLARNIFAT-----DSQNNWYFYG-SDGVAVTGS	1334
QY	1339	AKVDGRDLYFNPDGSGVQKGDFAFDBSGNTSFYHGDNGDKVVGQFFFTGNNMAYADNNG	1398
Db	1335	QTIAGKLYFASD-GKQVKGSFVT-YNGKVHYHHADSGELQVNRFEADKQGNWYLYDSNG	1392
QY	1399	NLVKGFQEI:DGKWHFBEVYTGQAKGAALVNGQOLYFDVDSGTVQKGDFTWD	1450
Db	1393	EALTGSQRI-----NDQRVFF-TREGQVKQGDVAYD	1422

RESULT 3

US-09-740-274-6

Sequence 6, Application US/09740274

; Patent No. 6465203

GENERAL INFORMATION:

APPLICANT: Nichols, Scott E.

TITLE OF INVENTION: Glucan-containing Compositions and Paper

Db 924 KVMADVPDMYAFPEKEVWVTRVVDKYGTPVAGSQIKNTLYVVDGKSSGKQQAQYCGA 983
QY 990 FLDELKAKYPDMFTNMISTGKPIDPSTKIKOWEAKYFNGTNTVLGKAGYVLSDATGKY 1049
Db 984 FLEELQAKYPELFAKQIISTGVPMDBSVKIKWSAKYFNGTILGRGAGYVLKQQAATNY 1043
QY 1050 FTVNENG--FLPASFTGONAKTGYDGTGMAYYSTSGNKAVNSFIYEGGHYFFDKD 1107
Db 1044 FNISDNKEINFLPKILL--NQDSQVGSYDGKGYVYSTSGYQAKNTFISEGDKWYFFDN 1102
QY 1108 GHMVTGSYKAEQNDYFLPENGIQMRDAIYQDAQNSVYVYGRITGLYKGDNVPFFVDPNN 1167
Db 1103 GWTGCA--QINGVNYFLUSGLQLDALKNEGTYIYNGDGRYE--NGYQFM---- 1156
QY 1168 ANKTVFRYPDANNVMAIGYRNMYGQTYYPDENGFOAKGOLLTDDKG--THYFDEDRGAMAK 1226
Db 1157 --SGVVRHEN--NGEMSVGLTVIDGVQVQYFDEMGYQAKGFVTTADGKIRYFDKQSGNMYR 1213
QY 1227 NKFEV--NVGDDWYMDGNGNAVKQYPPVNNQILYFNPETGVQVKGQFIDTAQGRTSYYDAN 1285
Db 1214 NREFIENEKGWLYLGBDGAATVGSQTINGQHLIFR--ANGVQVKGFEVTDHGRISYYDGN 1272
QY 1286 SGALKSSGFPTENGSDWYVAENGYYVYKGFQVAENQODWYFFDQTTGKQAKGAAGVDGRD 1345
Db 1273 SGQOINRF-----VRNAQGWFFYD--NNGYAVTGARTINGQL 1309
QY 1346 LYFNPDSGVQVKGDFATDESNGTSFYHGDNBGRKVGWGGFTTGNNAWYADNNGNLVKGFQ 1405
Db 1310 LYFRA--NGVQVKGFEVTDYGRISYYDNGSGQIRNRFVRNAQGWFFYDNNGYAV---- 1364
QY 1406 EIDGKWHFDEVTGQAKGAALVNGQOLYFDVDSGLQVKGDFVTDGQNTSYDYVNSGDK 1465
Db 1365 -----TGARTINGQHLIFRA--NGVQVKGFEVTDHGRISYYDNGSGDQ 1406
QY 1466 KVNQFPPTGDNAWYADGQNLAKGRKSIDNODLYFDPATGKQVKGOLYSID--GRNYFYD 1524
Db 1407 IRNRFVRNAQGWFFYDNNGYAVTGARTINGQHLIF--RANGVQVKGFEVTDYGRISYYD 1465
QY 1525 SGSGNMAK 1532
Db 1466 ANSGERV 1473

RESULT 6

US-09-740-274-2
; Sequence 2, Application US/09740274
; Patent No. 6455203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans

US-09-740-274-2

Query Match 41.7%; Score 3434.5; DB 2; Length 1475;
Best Local Similarity 46.6%; Pred. No. 9.1e-200;
Matches 731; Conservative 248; Mismatches 458; Indels 131; Gaps 39;
QY 1 MEKKLHYKLHKYKGGHWTIAVAS--LGLYSLVCAGTVSAEDKYANDTTAAQTAVGVDTGQDQ 59
Db 1 MDKRYRYKLKRYKGGWTVSVASAVMTLTLSGGLVKAD---SNESKSO----- 46
QY 60 ATTNDANTNTTDDTADQSAN-----TNODOAGSQDSNNQDQAKQDTAN----- 103
Db 47 -ISDNTSVV---TANESNVITEATSKQEAASSQTNHTVTTSSTSVVNPKEVSNP 102
QY 104 -TDRNQANSQTDNNQATDQATSPATDGTSVORRDAANVATAADQEGQTPASEQEKSAA 162
Db 103 YTVGETASNGEKLQNTT-----TVDKTS---EAAANNISKQTTTEADTDVIDDSSNAANL 153
QY 163 ----SLDNVKLIDGKYVYVQADGSYKKNFAITVNGOMLYFDSDTGALSSTSYTSFSG--T 217
Db 154 QILEKLPNVKEIDGKYVYDNNNGKVRTNFTLTADGKILHFD--ETGAYTDTSDTWNKOIV 212
QY 218 TNLVDDFSSHNAKAYDSTAKSFELVNGYLTANSWYRPAGILRNGQTWAEASNENDLRPLMS 277
Db 213 TTRSNIYKYNQVYDRSAQSFEHVDHYLTAESWYRPKYLKDGKTWTQSTEXDFRPLMT 272
QY 278 WMPDKDTQVAYVYNNKYLSANETEVNTNETSQVDLNKEAQSIOTKIEQKITSNDSTQWL 337
Db 273 WMPDQETQRYVYNNWALGINKT--YDDTSNQLNLNIAAATIAQAKIEAKITTLKNTDML 331
QY 338 TAMEAFVAAQPKWNMSTEN--FNKGDHLOGGALLYTN--SDLTTPWANSDYELLNRTPTQDG 395
Db 332 QTFISAFVKTQSAWNSDSEKPF--DHLQNGAVLYDNEGKLTTPYANSVRILNRTPTNQG 389
QY 396 TK--KYFTEGGGGYEFLLSDVDNSNPVQAEQLNQLHYLMNWGDIVMGDXDANPDGVR 453
Db 390 KDPRTADNTIGGYEFLLANDVDNSNPVQAEQLNLWFLNFGNIYANDPDANFDSIR 449
QY 454 VDAVDNVADLLQVYSNYFKONYKVTYDSEANALAHISILEAWSLNDNQNEBTNGTALSI 513
Db 450 VDAVDNVADLLQIAGDYLKAAKGIHKNDKAANDHLSILEAWSNDNTPYLHDDGDNMINM 509
QY 514 DNSSRLTSLAVLTQPGORIDLNLISSESVNKERANDTAYGDTIPTYSFVRAHDSVQTV 573
Db 510 DNKLKLLFLSLAKPLNQRSGNPLITNSL--VNRITDDNABTAAPVSPYSPIRAHDSVQDL 568
QY 574 IAKIVKEIDTNSDGYTFTLDOLDAFKIYNEDMAKVNKYTYTHYINIPAAAYALLLSNMSV 633
Db 569 IADIIKAEINPNVGYSTMEIEIKAFELYNKDLATEKKYTHYNTALSYALLTNKSSV 628
QY 634 PRVYDGYDLYDDQYMAKSPYDYDAIATMLQRIAYVSGGQSEEVHKNVNGNQILSSVRY 693
Db 629 PRVYDGMFTDDQYMAHKTINYEAITLLKARIKYVSGGQAMRNQOV--GNSIEITSVRY 687
QY 694 GQDMSADDTQCTDLSTRSGLVTLVSNQPNLDL--GQDSLTVNMGRANQVRPLILGK 752
Db 688 GKGALKATDT--GDRITRTSGNAVIEGNPNPSLKLKASDRVVVNMGAHKNQVRPLLLTDD 746
QY 753 DGVSQVLYKDSO--TNIVKYVTDANGNLTFTADDIKGYSTVDMSGYLAVVVPVGAK--DGQDV 809
Db 747 NGIKAYHSDQEAAGLVRYTNDREGELIFTAADIKGANPOVSGYLVGVVVPVGAALIKWFAL 806
QY 810 RVAADTNQADGKSLKTSAAALDSQVIEGFSNFQFANNADYTNKKIAENADFFKKLGI 869
Db 807 RLARPHQQMA---SVHQNAALDSRVMEGFSNFQAFATKKEEYTNVVIKRVNDPKFAEWGV 863
QY 870 TSFEMAPQVVSATDGSFLDSIIIONGYAFSDRVDLMSKNKYGSKDDLANKALKALHANGI 929
Db 864 TDFEMAPQVVSSTDGSFLDSVIQNGYAFTRDYDLGISKPNKYGTADDLKALKALHSKI 923
QY 930 QAIADWVPDQIYOLEGEEVWTAKRYSYGNPTFDAYINNALLYATNTKSSGSDYQAOYGA 989
Db 924 KVMADVPDMYAFPEKEVWVTRVVDKYGTPVAGSQIKNTLYVVDGKSSGKQQAQYCGA 983


```
QY 731 LTVNMGRAHANOAYRPLILGTGDKGVQSYLKDSB--TNIVKYTDANGNLFTTADDIKGYST 788
Db 832 LVVNGAAHKNQAYRPLLSKSTGLATYLDKSDVPAGLVRYTDNQNLTFTTADDIAGHST 891
QY 789 VDMGYLAVVVPVGAQGDQVAAADTNQADKSLKTSALSDSOVIVYEGFSNFQDFANN 848
Db 892 VEVGYLAVVVPVGAQGDQVAAADTNQADKSLKTSALSDSOVIVYEGFSNFQDFVKT 950
QY 849 DADYTNKXIAENADFFKLGITSFEMAPQVYSATDGSFLDSIIQNGYAFSDRYDLAMSKN 908
Db 951 PSQYTNRVIAQNAKLFKEWGITSFEPAPQVYSSQDGTFLDSIIENGAFEDRYDIAMSKN 1010
QY 909 NKYSGKDLALNALKALHANGIOAIAADVPDQIYQLPGBEVVTAKTNSYGNPTFDAYINN 968
Db 1011 NKYSGKDLALNALKALHANGIOAIAADVPDQIYQLPGBEVVTAKTNSYGNPTFDAYINN 1070
QY 969 ALYATNTKSSGSDYOAOYGGAFDELKAKYPMFTVMNISTCKPDPSTKIKOWEAKYFN 1028
Db 1071 SLYAAKTRTFGNDFOGKYGGAFDELKAKYPAIFERVQISNGRKLTTNEKITQWSAKYFN 1130
QY 1029 GTNVLGKAGYVLSDDATGKYFTVNEGDFLPASFTGQNAKTGFYDGTGMAYYSTGN 1088
Db 1131 GSNIGTGARYVLDNATNQYFSAKAGOTFLPKQMT--EITGSGFRFVGGDDVQVLSIGY 1188
QY 1089 KAVNSFIYEGGH--YYFDKGDHMTGSKYAEKNDYIFLPGIOMRDALYQDAQNSYY 1147
Db 1189 LAKNTFIQVGAQWYFYDKNGNMVTGE--QVIDGKYFFLDNGLQLRHVLRQSGDGHVY 1247
QY 1148 GRTGI--LYKGDWNPVDPNNANKTVFVFDANNVMAIGYRNMVGTYYFDS--NGFOAKG 1205
Db 1248 DPKGVQAFNG--FYDPAGB--RODVFYDNGQMYRGLHDMYGTTFYFDEKTLGIAQD 1301
QY 1206 QLLT--DDKGTHTYFEDENGAMAKNKFVNVGDD--WYMDGNGNAVKGQVPPVNNQILYFNP 1262
Db 1302 KFIRPADGRTYFIPDTGNLAVNRPAQNPENKAWYLDNSGYAVTGLQTINGKQYFDNE 1361
QY 1263 TGQVQKQFIPDAQRTSYDANSALKSSGFFTPNGSDWYAEANGYVYKGFQVAENQD 1322
Db 1362 -GRQVKGHEVT--INNQRVFLDGSGEIAPSRFVTEN-----N 1396
QY 1323 QWYVFDQTTGKAAGKAVDGRDLVFNPDGSGVQVKGDPATDESGNTSFYHGDNGDKWVG 1382
Db 1397 KWIYVD--GNGKLVGAQVINGNHYYFNNDYS--QVKGAWA-----NGRYVDGSGQAVSNQ 1449
QY 1383 FTTTGNNAWYADNNGNLKGFQEIIDGKWYHFDEVDTGQQAAGALVNGQQLYFDVDSGIQ 1442
Db 1450 FIQIAANQWYALNQDGHKVTGLQINNNKYVFG-----SNGAQ 1487
QY 1443 VKGDFVTDGQNTSYDVSNGDKKNGPFTTGDNAWYADGQGNLAKGRKSIDNQDLYPD 1502
Db 1488 VKGKLLT--VQKKCYFDAHTGEGVNVRFYEAARGCWYFNSAGQAVTGQOVINGKQLYPD 1546
QY 1503 PATGQVKGQVLSIDGRNYPFSDSGSNMAKNR 1534
Db 1547 -GSGRQVKGRYVYVGGKRLFCDAKTYGELQR 1577
```

RESULT 8

```
US-09-210-361-4
; Sequence 4, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
```

```
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
; US-09-210-361-4
```

```
Query Match 38.4%; Score 3161.5; DB 2; Length 1375;
Best Local Similarity 48.5%; Pred. No. 3.1e-183;
Matches 662; Conservative 223; Mismatches 400; Indels 81; Gaps 31;
```

```
QY 1 MEKKHLKHLKHKHWTIIVAS--IGLVSLVGA--GTVSAEDKVANDTTAQTAVGVDTG 56
Db 1 MEKKVRFKLRKVKRWVTVSIAAVVTLTSLGSLVKADSTDDRQQAQVTEQASL----- 55
QY 57 QDQATTNDANTNT--TDDTADQSANNTNODQAG--SQSNNOQAKQDUTANT-----DRNQ 108
Db 56 ---VTTSEAAKETLTATDTSTATSATSOPTATVDNVSTTNTSTNTTANTANFVVKPTTT 112
QY 109 ADNSOTDN-----NQATDQATS-----PATDGTSVQRRDAANVATAADQEG---QTAP 153
Db 113 SEQAKTDSNDKIIITTSKAVNRLTATGKVPANNNTAHPKVTDTKIVPIKPKIGLKQPSS 172
QY 154 SEQEKSAAL--SLDNVKLIDGKYVYVQADGSKYKKNFAITVNGQMLYFSDTGTALSTSTYS 212
Db 173 LSQDDIAALGNVKNIRKVGKYYKYKEDGTLQKNYALNKGKTFPFD--ETGALSNTLPS 231
QY 213 FSQGTIN--LVYDFFSHNKAYDSTAKSPELVNGYLITANSWYRPAGILRNGQWTEASNND 270
Db 232 KKGNTITNDNTNSFAQYNQVSTVDVANFEVHDYHLYTAESWYRPKYLKDKGKTWTQSTEKD 291
QY 271 LRPVLMSWMPDKQTVAYVYNNMKNYLSANETEVNTNETSQVDLNKEAQSIQTIKIBQKITS 330
Db 292 FRPLMTWMPDQETQRYVYNNMNAQLGHTQVNT--ATSPQLNLAAQTITQKIEKITAE 350
QY 331 NSTQWLRAMEAFVAAQPKWNNSTEN--FNKGDLHOGGALLYT--NSDLTPWANSDFLLNR 388
Db 351 KNTNWLRTTISAFVKTQSAMNSDSEKPF--DHLQKALLYSNNKSLTSQANSRYILNR 408
QY 389 TPTQDQGTG--KYPTGEGGEGVEFLSNDVNSNPVQAEOLNQLHLYLMNWGDI VMGDKD 446
Db 409 TPTNQTGKDPRTYADRTIGGYEFLLANDVNSNPVQAEQLNWLHFLMNFNIYANDPD 468
QY 447 ANFDGVRVDAVDNVDNADLLQVYNSYFNKDYKVTDSANALAHISILEAWSLNDNQYNEPT 506
Db 469 ANFDSIRVADVNDVDADLLQIAGDYLKAAGIKHKNDKKAANDHLSILEAWSYNDPTYLHDD 528
QY 507 NGTALSIDNSRLTSVLAVITTPQGORIDLNLISSEVKNKERANDTAYGDTITTYTSFVRAH 566
Db 529 GDNMINMDNRLRLSLYSLSLAKPLNQSGMNPILTNSL--VNRITDDNAETAAPSYSFIRAH 587
QY 567 DSEVQTVIAKIVKEKIDTNSDGYTFTLQDLKDAFKIYNEDMAKVNKTYTHYIPAYALL 626
Db 588 DSEVDQLIRNIIRTEINPNVVGYSFTTEIKKAFIYNKDLATATEKKYTHYITALSYALL 647
QY 627 LSNMESVPRVYVYDLYTDDGQTMAKKSPYDAIATMLQRIAYVSGGQSEEVHKVNGNNO 686
Db 648 LTNKSSVPRVYVYDLYTDDGQTMAKKSPYDAIATMLQRIAYVSGGQAMRQOV--GNSB 706
QY 687 ILSSVRYGODLMSADDTQGTDLSPRTSLVTLVSNDPNLDL--CGDSLTVNMGRAHQAQR 745
Db 707 IITSVRYGKGAUKATDT--GDRTRTSGVAVIEGNNPFLKASDRVVNVNMGAAHKNQAYR 765
QY 746 PLILGTGDKGVQSYLKDSB--TNIVKYTDANGNLFTTADDIKGYSTVDMSGYLAWVVPVGA 804
```

```
Db 766 PLLLTTDNGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKGYANPOVSGYLVWVPVGA 825
QY 805 DGQDVRVAADTNOKADCKSLKTSAAALDSQVIYEGFSNFQDFANNDADYTNKKIAENADFF 864
Db 826 ADQDVRVAASTPSTDGKSHQNAALDSRVWVFSGFSNFQAFATKBEYTNVIAKNVDF 885
QY 865 KKLGTISFEMAPQYVSATDGSFLDSIIQNGYAFSDRYDLAMSNNKYKSGKDDLALAKAL 924
Db 886 AEWGVTDFEMAPQYVSSTDGSFLDSVQNGYAFSDRYDLAMSNNKYKSGKDDLALAKAL 945
QY 925 HANGIOAIADWVPDQIYOLPGEVVTAKRTNSYGNPTFFDAYINNALYATNTKSSGSDYQA 984
Db 946 HSKGIKVMADWVPDQIYALPEKEVVTATRVDTGTVPVAGSQIKNTLYVVDGKSSGKQQA 1005
QY 985 KYGGAFLDELKAKYPMFTNMISTGKPIDPSTKIKQWEAKYFNGTNVLGKAGYVLSDD 1044
Db 1006 KYGGAFLDELKAKYPMFTNMISTGKPIDPSTKIKQWEAKYFNGTNVLGKAGYVLSDD 1065
QY 1045 ATGKYTVNENGDFLPASFT----GDONAKTGPYDGTGMAYYSTSGNKAVNSFIYEGGH 1100
Db 1066 ATNTYFSLVSDNTFLPKSLVNPNGHGTSSVTGLVFDGKGYVYSTSGNQAKNAFISLGN 1125
QY 1101 YYYFDKGHMVTSYKAEQNDYVFLPNGIQMRDAIYQDAQNSYYVYGRGILYK----- 1155
Db 1126 WYFDFNNGYVMTGA-QSINGANYFLSNGIQLRNALYDNGNKVLSYVYNGDRYENGYYL 1184
QY 1156 -GDMWYFPVDPNANKTVFERYFDANNVMAIGYRNMVGTQVYDPDENGFOAKGOLLTDKG- 1213
Db 1185 FGQW-----RYFO-NGIMAVGLTRVHGAQVYFDASFOAKGQFITTADGK 1229
QY 1214 THYFDEBNGAMAKNFV-NVGDDWYVYDNGNNAVKGOYVNNQILYFNPETGVQVKGQFT 1272
Db 1230 LRYFDRDSNGQISNRFRVRSKGEWFLDHNGVAVTGTVTFGQRLYFKP-NGVQAKGEFI 1288
QY 1273 TDAQRTSYDANSALKSSGFTPNGSWY-YAENGYVYKGFQV 1317
Db 1289 RDANGYLRYDPSNGNEVRNFRVRSKGEWFLDHNGIATVGARV 1334
```

RESULT 9

```
US-09-740-274-4
; Sequence 4, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740, 274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutants
US-09-740-274-4
```

Query Match 38.4%; Score 3161.5; DB 2; Length 1375;

```
Best Local Similarity 48.5%; Pred. No. 3.1e-183;
Matches 662; Conservative 223; Mismatches 400; Indels 81; Gaps 31;
QY 1 MEKKLVKHKVKKHVVTTIAVAS--IGLVSLVGA--GTVSAEDKVAANDTTTAQATVGVDTG 56
Db 1 MEKKVFKLKKVKKVWTVTSIASAVTLTSLSSGLVKADSTDDRQQAQVTSQASL----- 55
QY 57 QDQATTNDANTNT-TDPTDQASANTNQDQAG-SDQSNNDQDQAKQDQANT-----DRNQ 108
Db 56 ---VTTSEAAKETLTATDTSTATSATSQPTATVTDNVSTTNTQSTNTTANTANFVVKPTTT 112
QY 109 ADNSQTDN-----NQATDQATS-----PATDGTSVQRDAANVATAADQEG---QTAP 153
Db 113 SEQAKTDNSDKIITTSKAVNRLTATGKFPANNNTAHPKVTVDKIPIKIKGLKQPS 172
QY 154 SEQEKSAAL-SLDNVKLIIDGKYVYVQADGSKKNFAITVNGOMLYPDSDFGALSSSTSTYS 212
Db 173 LSQDDTAAALGNVKNIRKNGKYYKYKEDGTQKNYALNNGKTFPFD-ETGALSNNLTLS 231
QY 213 FSQGTN--LVDDFFSHNKAYDSTAKSFELVNGYLTANSWYRPAQILRNGQITWEASNEND 270
Db 232 KKGNIITNDNTNSFAQYNQVYSTDVANFEVDHLYTAESWYRPKYILKDGKTTWTQSTEKD 291
QY 271 LRPVLSWVPDKDTQVAYVYNNKYLISANETEVNETSQVDLNEKASQISQTKIEQKITS 330
Db 292 FRPLMTWVPDQSTQRYVYNNNAQIGIHQTYNT-ATSPQLQLMLAAQTQTKIEEKITAE 350
QY 331 NSTOMLRTAMEAFVAAQPKNNKSTEN-FNKGDHLQGGALLYT-NSDLTPWANSDYRLNLR 388
Db 351 KNTNWLRLQITISAFVKTQSAWNSDSEKFPD--DHLQGLLYSNNKSLTQANSNRYILNR 408
QY 389 TPTQDQGTK--KYFTGEGGEGYFLLSNDVDSNPVVAEQNLMLHLMWGDIVMGDKD 446
Db 409 TPTNQTKGKPRYATADRTTIGGYEFLANDVDSNPVVAEQNLMLHLMWGDIVMGDKD 468
QY 447 ANFDGVRVDAVNVNADLLQVYNTFKYKVTDSANALAHISILEASLNDNQVNETD 506
Db 469 ANFDSIRVDAVNDVADLLQIAGDYLKAAKGHIKNDKAAANDHLSILEASYNPTPLHDD 528
QY 507 NGTALSIDNSRLTSLAVLTQPKQRI DLSNLISSEVKNKERANDTAYGDTIPTYSFVRAH 566
Db 529 GDNMNMNDRLSLLSYLAKPLNQRSGMNPILITSL-VNRTDDNAAETAAVPSYSIRAH 587
QY 567 DSEVQTVIAKIVKEKIDTNSDGYTFTLDQLDAFKIYNEDMAKVNKTYTHYINPAAYALL 626
Db 588 DSEVQDLIRNIIRTEINPNVGVYFTTEIKKAFIYNKDLATEKYYTHYNTALSYALL 647
QY 627 LSNMESVPRVYGDLYTDDGQYMAKSPYDAIATMLQGRIAYVSGGQSEEVHKVNGNQ 686
Db 648 LTNKSSVPRVYGDLYTDDGQYMAKHTINYEAIETLLKARIKYVSGQAMRNQV-GNSE 706
QY 687 ILSVRYGODLMSADDTQGTDL-SRTSGLVTLVSNQPNLQD-LGDSLTVMNGRAHQAAYR 745
Db 707 IITSVRYGKALKATDT-GDRTTRTSGVAIVIEGNPSPRLKASDRVVVNMGAHQAAYR 765
QY 746 PLILGTQGVQSYLKDSQ--TNIYKTYDANGNITFTADDIKGYSTVDMSGYLAWVVPVGA 804
Db 766 PLLLTTDNGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKGYANPOVSGYLVWVPVGA 825
QY 805 DGQDVRVAADTNOKADCKSLKTSAAALDSQVIYEGFSNFQDFANNDADYTNKKIAENADFF 864
Db 826 ADQDVRVAASTPSTDGKSHQNAALDSRVWVFSGFSNFQAFATKBEYTNVIAKNVDF 885
QY 865 KKLGTISFEMAPQYVSATDGSFLDSIIQNGYAFSDRYDLAMSNNKYKSGKDDLALAKAL 924
Db 886 AEWGVTDFEMAPQYVSSTDGSFLDSVQNGYAFSDRYDLAMSNNKYKSGKDDLALAKAL 945
QY 925 HANGIOAIADWVPDQIYOLPGEVVTAKRTNSYGNPTFFDAYINNALYATNTKSSGSDYQA 984
Db 946 HSKGIKVMADWVPDQIYALPEKEVVTATRVDTGTVPVAGSQIKNTLYVVDGKSSGKQQA 1005
QY 985 KYGGAFLDELKAKYPMFTNMISTGKPIDPSTKIKQWEAKYFNGTNVLGKAGYVLSDD 1044
```

Db 1006 KYGGAFLLELOAKYPELPARKQISTGVPMDPSVKIKQWAKYFNGTNIILGRGAGYVLKQ 1065
QY 1045 ATGKYFTVNEGDGFLPASTF---GDONAKTGYDGTGMAYYSTSGNKAIVNSFYEGGH 1100
Db 1066 ATWTFSLVSDNTFLPKSLVNPNGHTSSSVTGLVPDGKGVYVYSTSGNQAQNAFISLGN 1125
QY 1101 YYYFDKDHMTGYSYKABDGNDDYFLPNGIQMRDAIYQDAQNSYYVYGTGILYK----- 1155
Db 1126 WYFDNNGYMTGA-QSINGANYFLSNGIQLRNIYDNGNKVLSYVYDNGRRYENGYL 1184
QY 1156 -GDMWYFPFDPNNANKTVFRYFDANNVMAIGYRNMVYGTYYFDENGFOAKGOLLTDDKG- 1213
Db 1185 FGOOW-----RYFO-NGIMAVGLTRVHGAVQYFDPASGFOAKGQFIITADCK 1229
QY 1214 THYFDEBNGAMAKNFV-NVGDDWYMDGNGNAVGQYPVNNQILYFNPETGVQVKGQFI 1272
Db 1230 LRYFDRDSGNQISNRFPVRNSKGEWFLPDHNGVAVTGTVTFNGQRLYFKG-NGVQAKGEFI 1288
QY 1273 TDAQGRTSYDANSALKSSGFFTPNGSDWY-YAENGYYVYKFKQV 1317
Db 1289 RDANGYLRYYDPNSGNEVRNFRVRSKGEWFLPDHNGIATVGARV 1334

RESULT 10

US-09-499-203-2
; Sequence 2, Application US/09499203
; Patent No. 6570065
; GENERAL INFORMATION:
; APPLICANT: KOSSMANN, Jens
; APPLICANT: WELSH, Thomas
; APPLICANT: QUANZ, Martin
; APPLICANT: KNUTH, Karola
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternansucrase
; FILE REFERENCE: 147-196P
; CURRENT APPLICATION NUMBER: US/09/499,203
; CURRENT FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2057
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-09-499-203-2

Query Match 29.4%; Score 2422; DB 2; Length 2057;
Best Local Similarity 37.3%; Pred. No. 4.1e-138;
Matches 649; Conservative 220; Mismatches 577; Indels 294; Gaps 64;
QY 43 NDTTAQATVGVDTGQDQATNDANTTTDT-DADQSANTN-----ODQAGSD 89
Db 210 DSVTGKASNVDPVINGKAQGYDAQGNLKKSVYVADSSGQTYYPDNGQPLGLQTTIDGNL 269
QY 90 QSNQOD--QAK---QDANTDRNOADNSOTDNNQADQATSPATDGTQSVQRDAANVATA 144
Db 270 QYFNQGVQVKGQFQVNNKRIYFAPN--TGNNAVANTEILNGKQORDANGNOVKY-APS 326
QY 145 ADQEGTAPSEOFKSAALSDNVKLDGKYVYVQADGSKYKQNFPAITVNGQMLYFSDTGA 204
Db 327 KDVAAGTFFD---ANGVMTGLQITSGKTYLDEQHLKRNKYNAGTFNNQFMYFDADTGA 383
QY 205 LSTSTYSFSQGTNLVDFPSSH--KAYDSTAKSPELVNGYLTANSWYRPAGILNGOT 262
Db 384 GTAIBEQDQGLVDSQNSNTPHNAKSYDKS--SPENVGVLTADTWYRPTDILKNGDT 441
QY 263 WEASNENDLRPVLSMWPKDQTVAYVYVNMKYLKSAETVNETSQVDLNKEAQSIQTK 322
Db 442 WTASTEDNRPLMTWPKQTOANYLNPMSSKGLGITTYTAATSQKTLNDAAFPVQTA 501
QY 323 IEQKITSNDSOTWLTAMFAVAAQPKNNMSTEN--FNKGDHLQGGALLY-TNSDITPWA 379
Db 502 IEQKISLKSTEWLRDAIDSFVKTOANNKQTEDEAFDGLQWLQGGFLAYQDDSHRTPNT 561
QY 380 NS-DYELLNRTPTQDQTKKYFTTEGEGGYEFLLSNDVNSPNVQAEQLNLHLYLMWG 438

Db 562 DSGNRKLRQRPINIDGSKD--TTDGKS-EFLANDIDNSNPIVOAEQLNLHLYLMWG 618
QY 439 DIVMGDKDANFDGVRVADVNVNADLLQVYSYFNKDYKVTVDSEANALAHISLEAWSLN 498
Db 619 SITGNDNANFDGIRVADVNDVADLLKIAGDYFKALYGTGDKSDANANKHLSILEDWNGK 678
QY 499 DNQYNEDTNGTALSIDNSRLTSLAVLTQPCQORIDL-----SNLISESVNKR 547
Db 679 DPQYVNOQNAQLTWDYTVTSQFNSLTHGANRNSNMVFLDTGYLNGDLNKKIVDKNR 738
QY 548 ANDTAY-----GPT--LPTYSFVRANDSEVQTVIAK-----IVKEKIDTNSDGYTTF 592
Db 739 PMSGTLVNRANSQDKVIPNYSFVRAHDYDQDPIRKAMIDHGIKMQD-----TFT 792
QY 593 LDQLKDAFKIYNEDMAKVN--KTYTHYNIPIAAYALLLSNMESVPRVYVYGLTDDQYMA 650
Db 793 PDOLAQGMFEYVYKQDENPSPGPKKYNDYNLPSAYAMLLTNKDTVPRVYVYGMYLEGQYME 852
QY 651 KKSPIYDAIATMLQRIAYVSGQ-----SEVHKVNGNNOILSSVRYQDLMASDQTG 705
Db 853 KGTIYNPVISALLKARIKYVSGQTMATDSSGKDLKOGETDLTTSVRFKGIKMTSDQTTT 912
QY 706 TDLS---RTSGLVTLVSDNPNLDLGGD-SLTVNMGRAHANQAYRPLILGTGKGVQSYLKD 761
Db 913 QNSQDYKNOGGIGVIVGNPNDLKANDKTITLHMGAKHNQYRALVLSNDSGIDVYDSD 972
QY 762 SDTNIYKVTANGNL-----TFTADD-----IKYSTVDMSGYLAIVVVGAKDQD 808
Db 973 DKAPTLL-TNDNGDLI PHKNTTFVKQDGTI INYEMKSGLSNALISGLVGVVVPVGSQSD 1031
QY 809 VR-VAADTNQKADGKSLKTSALDSQVYIEGSPNFQDPAANDADYTNKKIAENADFFKL 867
Db 1032 ARTVATESSSSSNDGVSFHSNAALDSNVYIEGSPNFQAMPTSPQSNTVVIATKANLFEL 1091
QY 868 GITSFEMAPQVYSATDG-----SFLDSIIQCYAFSDRYDLAWSKN-----KYGSKDD 916
Db 1092 GITSFELAPQVRSRGSDTNGVSGMSFLDSFLNNGYATFTDYDLGFNKGADGNPNTKYGTD 1151
QY 917 LANALKALHANGIAIADWVPDQIYQLPGEVYVTAKTNSYGNPTFPDAYINNALYATNTK 976
Db 1152 LRNAIEALHKGQMAIADWVPDQIYALPGKEVVTATRVDERGNQLKDTDFVNLVYVANTK 1211
QY 977 SSGSDYQAYGGAFLDELKAKYPMFTVMISTGKPIDBPTKI KOWEAKYFNGTNNVLGK 1036
Db 1212 SSGVDYQAKYGGEGFLDKREEYPSLFKQNVSTGQPIDASTKI KOWSAKYMNGTNIHRG 1271
QY 1037 AGYVLSDDATGKYFTVNEGD--FLPASFTGDQNAKTGFYDGTGMAYYSTSGNKAIVNSFI 1095
Db 1272 AYVVLKDWATNQYFNAIKTNEVFLPLQLQ-NKDAQTGTFISDASGVKYYSISGYQAADTFI 1330
QY 1096 YEG-GHYTYFDKDHMTG-----SYKAEADGNDYFLPNGIQMRDAIYQDAQ 1142
Db 1331 EDGNGWYVYFDKGYMVRSSQGENPIRTVETSVNTRNGN-YVFMNPGVELRKGFDTNSG 1389
QY 1143 NSYTYGRTGILYKGNWTFVDPNNANKTVFRYPANNVMAIGYRNMVYGT--YFPDENG 1201
Db 1390 NYVYFDQGMVR--DKYINDANN-----FYHLNVGDTMSRGLFKFDSDTLQYFASNGV 1442
QY 1202 QAKGQLLTDKGT-HYFEDDNGAMAKNFVNVGDDWYV-----1239
Db 1443 QIKDSYAKDSKNGKYFDSATGNTDGTAKQATWDGNGYITITSDANNTIGVNTDYATYIT 1502
QY 1240 -----DGNNGNAVKGQYVNNQILYFNPETGVQVKGQF-----1271
Db 1503 SSLREDGLPANAPYGVVTKDQNGNDLKWQY-INHTKQYEGQ--VQVTRQYTDKSGVSN 1559
QY 1272 -IT-----DAQRTSYDANSALKSSGFF-----PNGSDWYAEANGVYVYKFKQV 1317
Db 1560 LITFAGGLQGLQRLWD--SRALTMTPTPKTMNQISFISYANRNDGLFLNAPYQVKGY-Q 1616
QY 1318 AENQDQWYVFDQTTCKQAKGAAKVGRDLVFNPDGSGVQVKGDFATDESQNTSFYHGDNG 1377

Db 1617 AGSNQ-YKQOVT---IAGVANVSGKDWLSISFNGTQY---WIDSOALNTNTHDMNQK 1669
Qy 1378 KVVGGFFTTGN-----NAVY-----YADNNGNLVKGFBIDGKWHFDE---VTG 1419
Db 1670 VFN-----TTSNLGLFLNAPYRPGYKLAGLAKYNN-----QTVTVSQYFDDQGTWMS 1721
Qy 1420 QAKGAALVNGQLYFDVDSGIVKGDFTVDGQNTSYDVNSGDKKVNKGFFTTGDNAWY 1479
Db 1722 Q-----VVLGGQTVVNDHALAQWQ---VSD---TDQQLVYVNSGRN-DGLFL---NAPY 1766
Qy 1480 YADG-----QGNLAKGRK-----SIDNQDLYEPATGKQVKGOLVSI 1516
Db 1767 RGGSQSLIGHTADYNGHGVQVTKGQDAYGAQWRLITLNNQQWVD---SRALSTTMOA 1823
Qy 1517 DGRNYFDS-----GSGNMAKNRFVRIGDQWI-YFGNDGAATNL 1554
Db 1824 MNDNMVNSSQRTDGLWLNAPYTMGAKWAGDTRSANGRYVHISKAYSNEVGNTYYLTNL 1883

RESULT 11
US-09-604-957-3
; Sequence 3, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1278
; TYPE: PRT
; ORGANISM: Lactobacillus reuteri
US-09-604-957-3

Query Match 27.7%; Score 2285; DB 2; Length 1278;
Best Local Similarity 40.3%; Pred. No. 4.3e-130;
Matches 510; Conservative 189; Mismatches 409; Indels 156; Gaps 31;
Qy 52 GVDTGQDQATNDANTNTTDTADQSANQDQ-----AGSDQSNQD----- 95
Db 85 GONTVTDQFSKNYA---TTGGNFDYVYKVNQVEFSGWHATNQSDKDSQWIIIVLNGKE 141
Qy 96 -----QAKQDTANTDRNOA-----DNSQTDNNQATDQATSPAT---DGTSVQRDA 138
Db 142 VKRQLVNDTKEGAGFRNDVYKWPAINENSSMSGFGI---ITLPTVKNENVLVHRFS 199
Qy 139 ANVATA-----ADQEGQTAP---SEQBSAALSLDNVLKIDGKYVYVQ-ADGSKKNFAIT 190
Db 200 NDVKTGEGNVYDFWSELMPVWDSFQKNGPLQFGLQTINGQQYIDPTTGPQRKNFLQ 259
Qy 191 VNGOMLYFSDTCALSSSTYSFSQGTNLVDFFSHNKAYSSTAKSFELVNGYLTANSW 250
Db 260 SGNMIIYFSDTGCVGTNLELOPAKGTVSSNEQYRNGNAAYSDDKSIENVNGYLTADTW 319
Qy 251 YRPAGILRNGQTWEASNDRLPVLMSWPKDQTVAYVYNNKY---LSANETEVNTE 307
Db 320 YRPQILKDGTTWDSKETDMDRILMWMPNTLUTQAYILNYKQHGNLLPSALPPFNADA 379
Qy 308 SQVDLNKEAQSIQTKIEQKITSDNSTQWLRTAMEAFVAAQPKWNMSTENFN-KGDHLOGG 366
Db 380 DPAELNHYSEIVQONEIEKRISETGNTDRLTLMDHFTVNNPMWNKDSNVNFSGIFQCG 439
Qy 367 ALLYTNSDLTWANSYRLNLTPTQ-QDGTKYFTTEGGEGGYFELLSNDVNSPNVQA 425
Db 440 FLKYNSDLTPYANSYRLLRMPINIKQTYR-----QGEFLLANDIDNSPVVQA 491

Qy 426 EQLNQLHYLMNWGDIYVMGDKDANFDGVRVDAVDNVDNADLIQVYSNTFKONYKYVTSBANA 485
Db 492 EQLNWLYYLLNFCGTITANNDQANFDSVRVDAVDNADLWNIADQYFNAAVGM-DSDAVS 550
Qy 486 LAHISILEAWSLNDNOYEDTNGTALSIDNSSRLTSLAVLTQPGORIDLSNLISSEVVK 545
Db 551 NKHINILEDNWHADPEYFNKIGNPQLTMDDTIK-NSLNHGLSDATNRWGLDAIVHOSL-A 608
Qy 546 ERANDTAYGDTIPTYSFVRAHDSVOTVIKEDITNSDGYTFTLDQLKDAFKLYNE 605
Db 609 DRENNSTENVIPNYSFVRAHDNNSDQIQNAIRDV--TGKDHTHTTFDEQKIDAYIQ 666
Qy 606 DMAKNVNTYTHNIPAYALLSNMESVPRVYVYDLYDDGQYMAKSPYDYDAIATMLQG 665
Db 667 DQNSTVKKNLYNIPASYAILLTKDTPRVYVYDLYDDGQYMEHQTRYDYLTLNLLKS 726
Qy 666 RIATYVSGGSESEVHKVNGNNOILLSSVRYQDLSADDTQGTDLRSRTSLGLVTLVNDPND 725
Db 727 RVKYVAGGQSMQTMVSGNNILTSVRYGKAMTATDT-GTDETRTQGGIGVVVSNTPNLK 785
Qy 726 LG-GDSLTVNMGRAHANOAYRPLILGTGQVOSYKXSDTNI VKYTDANGNLFTADD-- 782
Db 786 LGVNDKVLHMGAAHKNQYRAVLTTTGDVINTSDQAP-VAMTDENGDLVLSHNLV 844
Qy 783 -----IKGYSTVDMSGYLAWVVPVCAKDGQDVRVAADTNQKADGKSLKTSALDS 832
Db 845 VNGKEADTAVQYANPDVSGYLAWVVPVCAKDGQDVRVAADTNQKADGKSLKTSALDS 904
Qy 833 QVIEGFSNFQDPANNDADYTNKKIAENADFPKLGITSEFEMAPQVVSATDGSFLSIIQ 892
Db 905 NVIFEAFSNFVYTPTKESERANVRIAQNADFFASLGFTSFEMAPQVNSKDRFLDSTID 964
Qy 893 NGYAFSDRYDLAMSKNKYKSKDDLALNALKALHANGIOAIADWVPDIOIYOLPGEVVVTA 952
Db 965 NGYAFSDRYDLAMSKNKYKSKDDLALNALKALHANGIOAIADWVPDIOIYOLPGEVVVTA 1024
Qy 953 RTNSYGNPTFDAYINNALYATNTKSSGSDYQAOYGGAFDELKAKYPMFTVNMISTGKP 1012
Db 1025 RVDDRGNVWKDALINNLYVNT-IGGGEYQKYGAGFLDKLQKLYPEITFKQVSTGVA 1083
Qy 1013 IPSTKIKOWEAKYFNGTNVKGAGYVLSDDATKYFTV-NENGDFLPASFTGD-ONAK 1070
Db 1084 IDPSQKITESAKYFNGTNILHRGSGYVLSKADG-QOYVNLGTTTQKFLPQLTGEKKQGN 1142
Qy 1071 TGFYVDGTGMAY-YSTSGNKAVNSFIYEG-GHYVYFDKDHVMTVGSYKADGNDYVFLPN 1128
Db 1143 EGFVKGNDGNYFYDLAGNMVKNMTFEDSVGNWYFPDQDKMVENKH----- 1189
Qy 1129 GIQMRDAIYQDAQGSYYGRTGILYKGDNWPYFVDPNNANKTVFRYFDANNVMAIGYRN 1188
Db 1190 -----FVDVDS----- 1195
Qy 1189 MYGQ--TYVFDENGFOAKQLLTDKGTHTYFDEBNGAMAKNFVNVGDDWYMDGNNAV 1246
Db 1196 -YXEGTGYFFLKNXGVSFRGLVQTDNGTYFDF-NYGMVNRQNTINAGAMIYTLDEKGLI 1253
Qy 1247 KQY 1250
Db 1254 KASY 1257

RESULT 12
US-09-995-749A-2
; Sequence 2, Application US/09995749A
; Patent No. 6867026
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: BO43388-CIP

```

; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1781
; TYPE: PRT
; ORGANISM: Lactobacillus reuteri
US-09-995-749A-2

Query Match      27.7%; Score 2285; DB 2; Length 1781;
Best Local Similarity 40.3%; Pred. No. 6.9e-130;
Matches 510; Conservative 189; Mismatches 409; Indels 156; Gaps 31;

QY 52 GVDTGQDQATTNDANTTTDTTADQSANTNDQ-----AGSDQSNQD----- 95
DB 588 GDNVTVDQSKNYA-----TTGNGFDYKVNQNGVEFSGWHATNQSDKDSQWIIVLVNGKE 644
QY 96 -----QAKQDTANTDRNQA-----DNSQTDNNQATDQATSPAT---DGTSVQRDA 138
DB 645 VKRQLVNDIKKEGAAGFNNDVYKVNPAIENSSMSFGQI--ITLPVTKNENVQLVHRES 702
QY 139 ANVATA-----ADQEGQAP---SEQEKSAALSNDVKLIDGKYVYVQ-ADGSYKKNFAIT 190
DB 703 NDVKTGEGNVDFWSELMPVKDSFKQNGPLKQFGLQTINGQQYYIDPTTGPQRKNFLQ 762
QY 191 VNGQMLYFSDTCALGSTSYSPSQGTNLVDVDFSSHKNKAYDSTAKSFFELVNGVLTANSW 250
DB 763 SGNWLYFSDTGCVGTGNALELQFAKGTSSNEQVRNGNAAYSDDKSIENVNGYLTADTW 822
QY 251 YRPAGILRGQTMWASNEENDLRVLMSPDKDQVAVYVNMKY---LSANETEVNTE 307
DB 823 YRPQKILKQGTWTWDSKETDMRILMVWHPNTLITQAYLYNMYKHGNLPSALPFENADA 882
QY 308 SQVDLNKEAQSIQTKIEQKITSNSTOWLRTAMEAFVAAQPKVNMSTENFN-KGDHLQGG 366
DB 883 DPAELNHYSEIVQONIEKRISSETGNTDMLRTLMHDFVTNPNMKNDSNVNFSGIFOQGG 942
QY 367 ALLYTWSDITPANSYRLLNRPQO-QDCTKKYFTFEGGEGVEFFLLSNDVDNSNPVQA 425
DB 943 FLKYNSDUTPYANSYRLLGRPMINIKQOTYR-----GQEFFLLANDIDNSNPVQA 994
QY 426 EQLNQLHYLMNWGDI VMGDKDANFDGVRVDAVNVNADLLQVYSNYFKDNYKVTSEANA 485
DB 995 EQLNWLYYLLNFCITITANNDQANFSDSVRVDAPDNIDADLMNIAQDYFNAAYGM-DSDAVS 1053
QY 486 LAHISILEAWSLNDNQYNTDGTALSIDNSRLTSVLVTKQPGQORIDLNSLISSVNK 545
DB 1054 NKHINILEDWNHADPEYFNKIGNPQLTMDDTIK-NSLNHGLSDATNRWGLDAIVHQS-L-A 1111
QY 546 BRANDTAYGDTIPTYSFVRAHDSVQTVIAKIVKEKIDTNSDGYTFTLDQKDAFYNE 605
DB 1112 DRENNSTENVIPNYISFVRAHDSNQDQIQAIRDV--TKQYHTFTFEQKQIDAYIQ 1169
QY 606 DMAKVNKTYTHYNIAPAAYALLSNMESVPRVYVYDLYTDGQVMAKSPYDAIATMLOG 665
DB 1170 DQNSTVKYNLYNIPASYAILLTNKTDTIPRVYVYDLYTDGQVMEHQTRYDYTLTLNLS 1229
QY 666 RIAYVSGQSSEVHKVNGNQNILSSVRYGODLMSADDTGCTDLRSRTSGLVTLVSNPNLD 725
DB 1230 RVKYVAGGQSMQMSVGGNNILTSVRYGKGAMTATDT-GTDETRTQGGVGVVSNTPNLK 1288
QY 726 LG-GDSILTNMGGAHANAQVRPLILCTKQGVQSGLKSDNTNIVKYTDANGNLTFTADD-- 782
DB 1289 LGVNDKVLHMGAAHAKNQQAARAAVLTTTGDGINVTSDQCAP-VAMTDENGDLVLSHNLV 1347
QY 783 -----IKGYSTVDSMGYLAWVPVGAQDQGVRAADTNQKADGKSLTSAAKLS 832
DB 1348 VNGKEADTAVQGYANPDVSGYLAWVPVVGASDNQDARTAPSTKNSGNSAYRTNAAFD 1407
```

```

QY 833 QVIYEGFSNFDQFANNADYTNKKIAENADFEKKLGITSFEMAPQVVSATDSGLDSIIQ 892
DB 1408 NVIFEAFSNFVYPTPKESERANVRIAQNAFFASLGFTSFEMAPQVNSKDRTEFLDSTID 1467
QY 893 NGYAFSDRYDLAMSKNNKYGSKDDLANALKALHANGIOAIADWVPDQIYQLPGEVVTAK 952
DB 1468 NGYAFTRDYDLGMSBPNKYGTDEDLRNAIQALHAKGLQVMADWVPDQIYNLPGKEVATV 1527
QY 953 RTNSYGNPTFDAYINNALLYATNTKSGSDYQAOYGAFIDELKAKYPDMFTNMISTGKP 1012
DB 1528 RVDDRGNVWKDAIINNLYVVT-ICGGEYQKYGGAFLDKLQKLYPEIFTKKQVSTGVA 1586
QY 1013 IDPSTKIKQWEAKYENGNTNLGKGAGYVLSDDATGKYFTV-NENGDFLPASTGD-QNAK 1070
DB 1587 IDPSOKITEWSAKYNGTNIHRGSGYVLLKADG-QQYVNLGTTTKQFLPIQUTGEKKQGN 1645
QY 1071 TCFYVDGTGMAY-YSTSGNKAVNSFIYEG-GHYHYFDDKGHMTVGSYKAEDGNDYVFLPN 1128
DB 1646 EGFVKNDGNYFYDLAGNWNKNTFIEDSVGNWYFFDQDGKVENKH----- 1692
QY 1129 GIQMRDAIYQDAQGSYYYGRTGILYKGDNWPYFVDPNNANKTVFRYPDANNVMAIGYN 1188
DB 1693 -----FVDVDS----- 1698
QY 1189 MYGQ--TYYPDENGFOAKQOLLTDDKTHYFDEDNAGAKNFVNVGDDWYMDGNGNAV 1246
DB 1699 -YGEKGTYPFLKNGVSFRGLQVQTONGTYFFD-NYGMKVRNQTINAGAMIYTLDENGKLI 1756
QY 1247 KGQY 1250
DB 1757 KASY 1760
```

RESULT 13

```

US-09-604-957-4
; Sequence 4, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-604-957-4
```

```

Query Match      19.9%; Score 1643; DB 2; Length 545;
Best Local Similarity 59.7%; Pred. No. 1.2e-91;
Matches 328; Conservative 73; Mismatches 124; Indels 24; Gaps 8;

QY 411 LLSNDVDSNPVQAEQLNQLHYLMNWGDI VMGDKDANFDGVRVDAVNVNADLLQVYSN 470
DB 1 LLANDIDNSNPVQAEQLNQLHYLMNWGDI VMGDKDANFDGVRVDAVNVNADLLQIASD 60
QY 471 YPKDNYKVTDSANALAHISILEAWSLNDNQYNTDGTALSIDNSRLTSVLVTKQPG 530
DB 61 YLKAHGVGDKSEKNAINHLSILEAWSLNDNQYNTDGTALSIDNSRLTSVLVTKQPG 530
QY 531 QRIDLSN-----LISESVNKERANDTAYGDTIPTYSFVRAHDSVQTVIAKIVE 580
DB 120 EK-DASNKNKEIRSGLEPVTNLSIN-NRSAEGKNSERMANYIFIRAHDSVQTVIAKIIKA 177
```


Qy	581	KIDTNSDGYTFTLDOLKDAFKIYNEDMAKUNITYTHYNI	PAAYALLLSNWSVPRVYVG	640
Db	178	QINPKTDGLTFTLDELKQAFKIYNEDMRQAKKTYQSNIFTAYALMSNKDSTIRLLY	237	
Qy	641	LYTDDGQYMAKSPYYDAIATWLOGRIAVUSGQSEEVHKVNGNQ	-----ILSSVR	692
Db	238	MYSDGQYMAKSPYYDAIDTLKARIKYAAGQDMKITIYVEGDSKHMWDVTGVLTSVR	297	
Qy	693	YGQDIMSADDTQGTDLSTRSGLTVLTVSNDPNLIDLG	-GDSLTVNMGRAHANQAYRPLILGT	751
Db	298	YGTGANEATD-QGSEATKTQGMAVITSNNPSLKLNQNDKVI	VNMGAAHKNOEYRPLLLTT	356
Qy	752	KDGVSQYILKDSDT-NIVKYTDANGMLTFTADDIKGYSTVDMSGYLA	VVPVGAKDQDVR	810
Db	357	KDGLTSYTSDDAAKSLYRKTNDKGELVFEDASDIQGYLNPQVSGYLA	VVPVGASDNQDVR	416
Qy	811	VAADTNQKADGKSLKTSRALDSOVLTYEGFSNFQDFANNDADYTNNK	KIAENADFFKLIGIT	870
Db	417	VAASNKANATQGVYESSALDSQLIYEGFSNFQDFVTKDSYDTNNK	KIAQNVLFFKSWGVT	476
Qy	871	SFEMAPQVVSATDGSFLLDSIIQNGYAFDRDYDLAMSKNNKYGS	KDDLANALKALHANGTQ	930
Db	477	SFEMAPQVVSSEDSGFLDSIIQNGYAFEDRYDLAMSKNNKYGS	QODMINAVKALHKSGLQ	536
Qy	931	AIADWVPDQ	939	
Db	537	VIADWVPDQ	545	

```

RESULT 14
US-09-995-749A-10
; Sequence 10, Application US/09995749A
; Patent No. 6867026
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAGUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: B043388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-995-749A-10

```

Qy	641	LYTDDGQYMAKSPYYDAIATMLQGRIAIYVSGGSEEVHKVGNQ-----ILSSVR	692
Db	238	MYSDGQYMATKSPYYDAITLLKARIKAYAAAGQDMKITYYEGDKSHMDWDYTGVLTSVR	297
Qy	693	YGQDLMSADDTQGTDLSTSGLTVLVSNDPNLDLG-GDSLFTVMGRAHANQYRPLLGT	751
Db	298	YGTGANEATD-QGSEATKTQGMVITSNPSLKLNQNDKVIYNMGAHKNQYRPLLTT	356
Qy	752	KDGVCSYLKQSDT-NIVKYTDANGNLTPADDIKG-YSTVDMSGVLAVWVPVGAKGDQV	809
Db	357	KDGTSTYSDAAKSLRKTKNDKELVPDASDIQGLYLNPQVSG-LAVWVPVGASDNQDV	415
Qy	810	RVAADTNQKADGKSLKTSAALDSQVIYEGFNFQDFANNDDATYTKIAENADFPKKLGI	869
Db	416	RVAASNKANATQGVYESSSALDSQIIYEGFNFQDFVTKDSDYTNKKIAQNQLPFSWGV	475
Qy	870	TSPFEMAPQYVSATDGSFLDSIIQNGYAFSDRYDLAMSKNNKYGSKDDLAKALKHANGI	929
Db	476	TSPFEMAPQYVSSEDSGFLDSIIQNGYAFEDRYDLAMSKNNKYGSGQODMINAKLHKSGI	535
Qy	930	QAIADWVPDQ	939
Db	536	QVIADWVPDQ	545

RESULT 15
US-09-995-749A-11
; Sequence 11, Application US/09995749A
; Patent No. 6867026
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: B043388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 522
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-09-995-749A-11

Query Match	17.5%	Score 1442.5;	DB 2;	Length 522;
Best Local Similarity	56.1%;	Pred. No. 1.7e-79;		
Matches 298;	Conservative	76;	Mismatches 146;	Indels 11; Gaps 7
Qy	411	LLSNDVNSPPVQAEQLNHYLMWGGDILVMGDKDANFDFGVRVDADVNNADLLQVYSN	470	
Db	1	LLANDVNSNVVEARQLNWLYLMMFGTITANDADANFDFGIRVDADVNDVDRDLQLTARD	60	
Qy	471	YFKDNYKVTDSEANALAHISILSAWSLNDNQYNEDTNGTALSINSSRLTSLAVLTQPG	530	
Db	61	YPKLAYGVQNDATANQHLSILELDSWSHNDPLYVTDQGSNQLTMDYDVHTQLIWSLTSSD	120	
Qy	531	QRIDLNLSESUNKERANDTACDITPTYSFVRAHDSSEVQTVAKIVKEKIDTNSDGYT	590	
Db	121	IRGTMRQFV--DYVMVDRSNDSTENEAIPNYSFVRAHDSSEVQTVIAQVSDLYPDEVNSLA	179	
Qy	591	FTLDQLKDAFKIYNEDMAKVNKYTYHNI PAAYALLSNMESVPRVYGDLYTDDGYMA	650	
Db	180	PTTEQLAAAFKVNEDBKADKKYTYQNMAASAYAMLLTNKDTVPR--YGDLYTDDGGYMA	238	
Qy	651	KKSPYYDAIATMLQGRVAYVSGQSEVHVHKVNGNQLSSVRYGQDLWSADDDTGTDLSR	710	

Db 239 TKSPYDAINTLLKARVQVAGQMSV-----DSNDVLTSVRYGKDAMTASDT-GTSETR 293
QY 711 TSGLVTLVSNDFNLDL-GGDSLTVNMGRAHANQAYRPLILGTRKDGVSYLKSDTNI-VK 768
Db 294 TEGIGVIVSNAELQLEDGHTVTLHMGAHKKQAYRALLSTTADGLAYY--DTDENAPVA 351
QY 769 YTDANGNLFTTADDIKGYSTVDMGYLAVWVPVGAKDQDVRAADTNOKADGKSLKTS 828
Db 352 YTDANGDLIFTNESIYGVQNPQVSGYLAVWVPVGAQQDQDARTASDTTNTSDKV FHSNA 411
QY 829 ALDSQVIYEGFSNFOFANDADYTNKKIAENADPFKLGITSEFEMAPOYVSATDGSFLD 888
Db 412 ALDSQVIYEGFSNFOFATDSSEYTNVIAQNADQFKQMGVTSFQLAPQYRSSTDTSF 471
QY 889 SIIQNGYAFSDRYDLAMSNNKYGSKDDLANALKALHANGIOAIADWVPDQ 939
Db 472 SIIQNGYAFTRDYDLGYGTPTKYGTADQLRDAIKALHASGIOAIADWVPDQ 522

Search completed: February 11, 2006, 19:46:02
Job time : 42.0237 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 20:39:23 ; Search time 139.093 Seconds
(without alignments)
4668.162 Million cell updates/sec

Title: US-10-797-821-38

Perfect score: 8237

Sequence: 1 MEKKLHYKLHKVKKHWTTIA.....FVRIGDQWTFYFGNDGAATNL 1554

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA_Main:
1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pap.*
2: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pap.*
3: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pap.*
4: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pap.*
5: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pap.*
6: /cgn2_6/prodata/1/pubpaa/US11_PUBCOMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8237	100.0	1554	US-10-383-930-38	Sequence 38, Appl
2	8237	100.0	1554	US-10-797-821-38	Sequence 38, Appl
3	3818	46.4	1518	US-10-383-930-40	Sequence 40, Appl
4	3818	46.4	1518	US-10-797-821-40	Sequence 40, Appl
5	3615	43.9	1430	US-09-740-274-6	Sequence 6, Appli
6	3615	43.9	1430	US-10-383-930-36	Sequence 36, Appl
7	3615	43.9	1430	US-10-797-821-36	Sequence 36, Appl
8	3502.5	42.5	1590	US-10-383-930-37	Sequence 37, Appl
9	3502.5	42.5	1590	US-10-797-821-37	Sequence 37, Appl
10	3434.5	41.7	1475	US-09-740-274-2	Sequence 2, Appli
11	3434.5	41.7	1475	US-10-383-930-34	Sequence 34, Appl
12	3434.5	41.7	1475	US-10-797-821-34	Sequence 34, Appl
13	3161.5	38.4	1375	US-09-740-274-4	Sequence 4, Appli
14	3161.5	38.4	1375	US-10-383-930-35	Sequence 35, Appl
15	3161.5	38.4	1375	US-10-797-821-35	Sequence 35, Appl
16	3080.5	37.4	1365	US-10-383-930-39	Sequence 39, Appl
17	3080.5	37.4	1365	US-10-797-821-39	Sequence 39, Appl
18	2917	35.4	1497	US-10-484-218-18	Sequence 18, Appl
19	2822	34.2	1595	US-10-484-218-20	Sequence 20, Appl
20	2822	29.4	2057	US-10-417-280A-2	Sequence 2, Appli
21	2285.5	27.7	1777	US-10-484-218-12	Sequence 12, Appl
22	2285.5	27.7	1777	US-09-995-749A-2	Sequence 2, Appli
23	2256	27.4	1006	US-10-484-218-22	Sequence 22, Appl
24	2199	26.7	1771	US-10-484-218-14	Sequence 14, Appl
25	1615	19.6	545	US-09-995-749A-10	Sequence 10, Appl
26	1442.5	17.5	522	US-09-995-749A-11	Sequence 11, Appl
27	1277	15.5	787	US-10-484-218-16	Sequence 16, Appl

28	1265	15.4	535	3	US-09-995-749A-13	Sequence 13, Appl
29	1172.5	14.2	584	3	US-09-995-749A-12	Sequence 12, Appl
30	1158.5	14.1	525	5	US-10-484-218-23	Sequence 23, Appl
31	573	7.0	224	5	US-10-484-218-4	Sequence 4, Appli
32	556.5	6.8	223	5	US-10-484-218-10	Sequence 10, Appl
33	554.5	6.7	223	5	US-10-484-218-6	Sequence 6, Appli
34	545.5	6.6	2710	4	US-10-011-366-6	Sequence 6, Appli
35	545.5	6.6	2710	4	US-10-354-774-6	Sequence 6, Appli
36	545.5	6.6	2710	4	US-10-271-012-6	Sequence 6, Appli
37	545.5	6.6	2710	4	US-10-729-122-6	Sequence 6, Appli
38	545.5	6.6	2710	4	US-10-729-039-6	Sequence 6, Appli
39	545.5	6.6	2710	5	US-10-729-527-6	Sequence 6, Appli
40	545.5	6.6	2710	5	US-10-727-898-6	Sequence 6, Appli
41	545.5	6.6	2710	5	US-10-728-696-6	Sequence 6, Appli
42	545.5	6.6	2710	6	US-11-001-241-6	Sequence 6, Appli
43	510.5	6.2	221	5	US-10-484-218-8	Sequence 8, Appli
44	469	5.7	866	4	US-10-222-038-2	Sequence 2, Appli
45	469	5.7	866	5	US-10-987-038-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-10-383-930-38
; Sequence 38, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 38
; LENGTH: 1554
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-383-930-38

Query Match	100.0%	Score 8237;	DB 4;	Length 1554;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1554;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MEKKLHYKLHKVKKHWTTIAVASIGLVSLVGAGTYSAEKDVANDTTAQATVGVDTGDDQA	60	
Db	1	MEKKLHYKLHKVKKHWTTIAVASIGLVSLVGAGTYSAEKDVANDTTAQATVGVDTGDDQA	60	
QY	61	TTNDANTNTTDTADQSANNTQDQSGDSQNNQDQAKQDQTANTDRNQADNSQDNNQAT	120	
Db	61	TTNDANTNTTDTADQSANNTQDQSGDSQNNQDQAKQDQTANTDRNQADNSQDNNQAT	120	
QY	121	DOATSPATDGTSVQRDDAANVATAADQEQGTAPSEQESAAALSLDNVKLIDGKYVYVQAD	180	
Db	121	DOATSPATDGTSVQRDDAANVATAADQEQGTAPSEQESAAALSLDNVKLIDGKYVYVQAD	180	
QY	181	GSYKKNFATVNGQMLYFSDTGALSSSTYSFSQGTNNLVDDFSSHNKAYDSTAKSFEL	240	
Db	181	GSYKKNFATVNGQMLYFSDTGALSSSTYSFSQGTNNLVDDFSSHNKAYDSTAKSFEL	240	
QY	241	VNGYLTANSWTRPAGILRNGQTWEASNENDLRPLVMSWMPDKDQTVAYVYNNKYLISANE	300	
Db	241	VNGYLTANSWTRPAGILRNGQTWEASNENDLRPLVMSWMPDKDQTVAYVYNNKYLISANE	300	
QY	301	TEVTNETSQVDLNKEAQSIQTKIEQKITSNDSQMLRTAMEAFVAAQPKWNMSTENFNKG	360	
Db	301	TEVTNETSQVDLNKEAQSIQTKIEQKITSNDSQMLRTAMEAFVAAQPKWNMSTENFNKG	360	

QY	361	DHLQGGALLYTNSDLTPWANSYRLLNRTPTQODGTTKYFTTEGEGGYEFLLSNDVDSN	420
Db	361	DHLQGGALLYTNSDLTPWANSYRLLNRTPTQODGTTKYFTTEGEGGYEFLLSNDVDSN	420
QY	421	PVVQAEQLNQLHYLMNWGDI VMGDKDANFDGVRVDVAVNNADLLQVYNYFYKDNKYKVD	480
Db	421	PVVQAEQLNQLHYLMNWGDI VMGDKDANFDGVRVDVAVNNADLLQVYNYFYKDNKYKVD	480
QY	481	SEANALAHISILEAWSLNDQYNEEDTNGTALSIDNSSRLTSLAVLTQKPGQRIDLSNLS	540
Db	481	SEANALAHISILEAWSLNDQYNEEDTNGTALSIDNSSRLTSLAVLTQKPGQRIDLSNLS	540
QY	541	ESVNERANDTAYGDTIPYTSFVRAHDSEVQTVIKIVKEKIDTNSDGTFTFLDQLKDAF	600
Db	541	ESVNERANDTAYGDTIPTYSFVRAHDSEVQTVIKIVKEKIDTNSDGTFTFLDQLKDAF	600
QY	601	KIYNEDMAKNKTYTHYNIPAAVALLSNMESVPRVYVYGDLYTDDQYNAKKSPPYDATA	660
Db	601	KIYNEDMAKNKTYTHYNIPAAVALLSNMESVPRVYVYGDLYTDDQYNAKKSPPYDATA	660
QY	661	TMLQGRIAVYSGQSEEVHKVNGNNOILSSVRYGQDLSADDTQGTDLSTRTSLGLVLSN	720
Db	661	TMLQGRIAVYSGQSEEVHKVNGNNOILSSVRYGQDLSADDTQGTDLSTRTSLGLVLSN	720
QY	721	DPNLDLGGDSLTVNMGRAHANQAYRPLILGTGDKGVQSYLKSDSTNIVKYTDANGNLTFTA	780
Db	721	DPNLDLGGDSLTVNMGRAHANQAYRPLILGTGDKGVQSYLKSDSTNIVKYTDANGNLTFTA	780
QY	781	DDIKGYSTVDMGYLAVVPVGAQGDQVRVAADTNQKADGKSLKTSAAALDSQVIEGFS	840
Db	781	DDIKGYSTVDMGYLAVVPVGAQGDQVRVAADTNQKADGKSLKTSAAALDSQVIEGFS	840
QY	841	NFQDFANNADATYNNKIAENADPFKKLGITSEFMAPQYVSATDGSFLDSIIQNGYAFSDR	900
Db	841	NFQDFANNADATYNNKIAENADPFKKLGITSEFMAPQYVSATDGSFLDSIIQNGYAFSDR	900
QY	901	YDLAMSKNNKYGSKDILANALKALHANGIOAIADWVPDQIYQLPGBEVVTAKRTSYGNP	960
Db	901	YDLAMSKNNKYGSKDILANALKALHANGIOAIADWVPDQIYQLPGBEVVTAKRTSYGNP	960
QY	961	TPDAYINNALYATNTKSSGSDYQAOYGGAPLDELKAKYPDMFTVMNISTGKPIDPSKTK	1020
Db	961	TPDAYINNALYATNTKSSGSDYQAOYGGAPLDELKAKYPDMFTVMNISTGKPIDPSKTK	1020
QY	1021	QWEAKYFNGTGNVLGKAGYVLSDDATGKYFTVNEGDFLPASFTGQNAKTGYFDGTGM	1080
Db	1021	QWEAKYFNGTGNVLGKAGYVLSDDATGKYFTVNEGDFLPASFTGQNAKTGYFDGTGM	1080
QY	1081	AYYSTSGNKAVNSFIYEGGHYYFDKDGHMVTSYKABDGNYYFLPNGIQMRDAIYQDA	1140
Db	1081	AYYSTSGNKAVNSFIYEGGHYYFDKDGHMVTSYKABDGNYYFLPNGIQMRDAIYQDA	1140
QY	1141	QGNYYTYGRTGILYKGNWYPFVDPNNANKTVPRYFDANNVMAIGYRNMVGTQYTFDENG	1200
Db	1141	QGNYYTYGRTGILYKGNWYPFVDPNNANKTVPRYFDANNVMAIGYRNMVGTQYTFDENG	1200
QY	1201	FOAKGQLLTDKQTHYFDEDNAGAMAKNKFVNVGDDWYVMDGNGNAVKGQYPVNNQILYFN	1260
Db	1201	FOAKGQLLTDKQTHYFDEDNAGAMAKNKFVNVGDDWYVMDGNGNAVKGQYPVNNQILYFN	1260
QY	1261	PETGVQVKGQFITDAQRTSYDANSALKSGSFFFTPNGSDWYIYANGVYKFKQVAEN	1320
Db	1261	PETGVQVKGQFITDAQRTSYDANSALKSGSFFFTPNGSDWYIYANGVYKFKQVAEN	1320
QY	1321	QDQWYFDQTTGQAKGAACVGRDLYENPDSGVQVKGDFATDESNTSPYHGNDGKVV	1380
Db	1321	QDQWYFDQTTGQAKGAACVGRDLYENPDSGVQVKGDFATDESNTSPYHGNDGKVV	1380
QY	1381	GGFFTTGNNAWYADNNGNLVKGFQBEIDGKWHFDEVTGQAKGAALVNGQQLYFDVDSG	1440
Db	1381	GGFFTTGNNAWYADNNGNLVKGFQBEIDGKWHFDEVTGQAKGAALVNGQQLYFDVDSG	1440

QY	1441	IQVKGDFVTDGQNTSYDVNSGKKVNGFFTTGDNWYAYDQGNLAKGRKSIDNQDLY	1500
Db	1441	IQVKGDFVTDGQNTSYDVNSGKKVNGFFTTGDNWYAYDQGNLAKGRKSIDNQDLY	1500
QY	1501	FDPATGKQVKGQVLSIDGRNYYFDSGSGNMAKNRFRIGDQWIFYGNDGAATNL	1554
Db	1501	FDPATGKQVKGQVLSIDGRNYYFDSGSGNMAKNRFRIGDQWIFYGNDGAATNL	1554
RESULT 2			
US-10-797-821-38			
; Sequence 38, Application US/10797821			
; Publication No. US20050031633A1			
; GENERAL INFORMATION:			
; APPLICANT: Smith, Daniel J. .			
; APPLICANT: Taubman, Martin A.			
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens			
; FILE REFERENCE: 25669-020			
; CURRENT APPLICATION NUMBER: US/10797,821			
; CURRENT FILING DATE: 2004-03-09			
; PRIOR APPLICATION NUMBER: 10/383,930			
; PRIOR FILING DATE: 2003-03-07			
; PRIOR APPLICATION NUMBER: 60/363,209			
; PRIOR FILING DATE: 2002-03-07			
; PRIOR APPLICATION NUMBER: 60/402,483			
; PRIOR FILING DATE: 2002-08-08			
; PRIOR APPLICATION NUMBER: 09/290,049			
; PRIOR FILING DATE: 1999-04-12			
; PRIOR APPLICATION NUMBER: 60/081,550			
; PRIOR FILING DATE: 1998-04-13			
; PRIOR APPLICATION NUMBER: 60/115,142			
; PRIOR FILING DATE: 1999-01-08			
; NUMBER OF SEQ ID NOS: 45			
; SOFTWARE: PatentIn version 3.2			
; SEQ ID NO 38			
; LENGTH: 1554			
; TYPE: PRG			
; ORGANISM: Streptococcus sobrinus			
US-10-797-821-38			
Query Match 100.0%; Score 8237; DB 5; Length 1554;			
Best Local Similarity 100.0%; Pred. No. 0;			
Matches 1554; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	1	MEKKLHYKLHKVKKHWTIIVASIGLVSLVGAGTVAEDKVANDTTAQTATGVDTCQDQA	60
Db	1	MEKKLHYKLHKVKKHWTIIVASIGLVSLVGAGTVAEDKVANDTTAQTATGVDTCQDQA	60
QY	61	TTNDANTNTTDTADQSAANTNODQAGSDQSNNOQAKODTANTDRNOADNSQTDNNQAT	120
Db	61	TTNDANTNTTDTADQSAANTNODQAGSDQSNNOQAKODTANTDRNOADNSQTDNNQAT	120
QY	121	DOATSPATDGTSVQRDAANVATAADQEGQTAPSEQEKSAALSLDNVKLDIGKYIYVQAD	180
Db	121	DOATSPATDGTSVQRDAANVATAADQEGQTAPSEQEKSAALSLDNVKLDIGKYIYVQAD	180
QY	181	GSYKKNFAITVNGQMLYFDSGTGALSSTSTYSFSQGTTLNVDDFSSHNAKAYDSTAKSFEL	240
Db	181	GSYKKNFAITVNGQMLYFDSGTGALSSTSTYSFSQGTTLNVDDFSSHNAKAYDSTAKSFEL	240
QY	241	VNGYLTANSWYBPAGTLRNGQTVWEASNENDLRPVLMSWMPDKDQTVAYVNMNKLISANE	300
Db	241	VNGYLTANSWYBPAGTLRNGQTVWEASNENDLRPVLMSWMPDKDQTVAYVNMNKLISANE	300
QY	301	TEVTNETSQVDLNKEAQSITQTKIEQKITSNDSTQWLRTAMEAFVAAQPKNMSTENFNKG	360
Db	301	TEVTNETSQVDLNKEAQSITQTKIEQKITSNDSTQWLRTAMEAFVAAQPKNMSTENFNKG	360
QY	361	DHLQGGALLYTNSDLTPWANSYRLLNRTPTQODGTTKYFTTEGEGGYEFLLSNDVDSN	420
Db	361	DHLQGGALLYTNSDLTPWANSYRLLNRTPTQODGTTKYFTTEGEGGYEFLLSNDVDSN	420
QY	421	PVVQAEQLNQLHYLMNWGDI VMGDKDANFDGVRVDVAVNNADLLQVYNYFYKDNKYKVD	480

Db 421 PVVQAEQNLHLYLNNWGDIVMGDKDANFDGVRVDAVDNWNADLLQVSYNFYKDNKYKTD 480
QY 481 SEANALAHISILEAWSLNDQYNEDTNGTALSIDNSSRLTSLAVLTQPGQRIDLSNLIS 540
Db 481 SEANALAHISILEAWSLNDQYNEDTNGTALSIDNSSRLTSLAVLTQPGQRIDLSNLIS 540
QY 541 ESNVKEKANDTAYGDTPTTYSFVRAHSEVQTVIAKIVKEDIKIDNSDGYTFTLDQLKDAF 600
Db 541 ESNVKEKANDTAYGDTPTTYSFVRAHSEVQTVIAKIVKEDIKIDNSDGYTFTLDQLKDAF 600
QY 601 KYNEDMAKNKVTYTHNIPAAVALLSNMESVPRVYGYDLYTDDGOYMAKSPYDAIA 660
Db 601 KYNEDMAKNKVTYTHNIPAAVALLSNMESVPRVYGYDLYTDDGOYMAKSPYDAIA 660
QY 661 TMLQGRYAYVSGGSEEVHVKVNGNQILSSVRYGQDLMSADDTQGTDLSTSGTLVLSN 720
Db 661 TMLQGRYAYVSGGSEEVHVKVNGNQILSSVRYGQDLMSADDTQGTDLSTSGTLVLSN 720
QY 721 DPNLDLGGDSLTVNMGRAHQAQYRPLILGTQGVQSYLKSDSTNIIVKYTDANGNLTFFTA 780
Db 721 DPNLDLGGDSLTVNMGRAHQAQYRPLILGTQGVQSYLKSDSTNIIVKYTDANGNLTFFTA 780
QY 781 DDIKGYSTVDMSGYLAVWVPVGAQGDQVRAADTNQKADGKSLKTSAAALDSQVIYEGFS 840
Db 781 DDIKGYSTVDMSGYLAVWVPVGAQGDQVRAADTNQKADGKSLKTSAAALDSQVIYEGFS 840
QY 841 NFQDFANNADATYNNKIAENADPFKLGITSPFMAPOYVSATDGSFLDSIIQNGYAFSDR 900
Db 841 NFQDFANNADATYNNKIAENADPFKLGITSPFMAPOYVSATDGSFLDSIIQNGYAFSDR 900
QY 901 YDLAMSKNNKYGSKDDLALNALKALHANGIOAIADWVPDQIYQLPGBEVTAKRTNSYGNP 960
Db 901 YDLAMSKNNKYGSKDDLALNALKALHANGIOAIADWVPDQIYQLPGBEVTAKRTNSYGNP 960
QY 961 TPDAYLNALYATNTKSSGSDYQAOYCGAFDELKAKYPMFTVMNISTGKPIDPSTKIK 1020
Db 961 TPDAYLNALYATNTKSSGSDYQAOYCGAFDELKAKYPMFTVMNISTGKPIDPSTKIK 1020
QY 1021 QWEAKYFNGTNVLGKAGYVLSDDATGKFTVNNEDGDFLPASFTGQDNAGTGYYPDGTGM 1080
Db 1021 QWEAKYFNGTNVLGKAGYVLSDDATGKFTVNNEDGDFLPASFTGQDNAGTGYYPDGTGM 1080
QY 1081 AYYSTSGNAKVSFIYEGGHYYFDKGMVTVSYKAECDNDYFPLNGIQMRDAIYQDA 1140
Db 1081 AYYSTSGNAKVSFIYEGGHYYFDKGMVTVSYKAECDNDYFPLNGIQMRDAIYQDA 1140
QY 1141 QGNSYYGRTGILYKGDNHYPPVDPNNAKTVPRYFDANNVMAIGYRNMVGTYYPDENG 1200
Db 1141 QGNSYYGRTGILYKGDNHYPPVDPNNAKTVPRYFDANNVMAIGYRNMVGTYYPDENG 1200
QY 1201 FOAKGQLLTDKGTHTYFDEBNGAMAKNFVNVGDDWYMDGNGNAVKGOYPVNNQILLYFN 1260
Db 1201 FOAKGQLLTDKGTHTYFDEBNGAMAKNFVNVGDDWYMDGNGNAVKGOYPVNNQILLYFN 1260
QY 1261 PETGVQVKQFITDAQRTSYDANSALKSSGFFTPNGSDWYAEVGYVYKFKQVAEN 1320
Db 1261 PETGVQVKQFITDAQRTSYDANSALKSSGFFTPNGSDWYAEVGYVYKFKQVAEN 1320
QY 1321 QDQWYFDQTTGQAKGAQKVDGRDLYFNPDSGVQVKGFATDESNTSFYHGDNGDKVY 1380
Db 1321 QDQWYFDQTTGQAKGAQKVDGRDLYFNPDSGVQVKGFATDESNTSFYHGDNGDKVY 1380
QY 1381 GGFFTTGNNAWYADNNGNLVKGFBIDGKWHFDEVTQQAALVNGQOYLFDVDSG 1440
Db 1381 GGFFTTGNNAWYADNNGNLVKGFBIDGKWHFDEVTQQAALVNGQOYLFDVDSG 1440
QY 1441 IQVGDFVTDGQNTSYDVNSGDKKVGFFTTGDNANWYADGQGNLAKGRKSIDNQDLY 1500
Db 1441 IQVGDFVTDGQNTSYDVNSGDKKVGFFTTGDNANWYADGQGNLAKGRKSIDNQDLY 1500
QY 1501 FDPATGKQVKQLVSDGRNYFYDSSGSGNMAKNRFRVIGDQWIFYGNDGAATNL 1554

Db 1501 FDPATGKQVKQLVSDGRNYFYDSSGSGNMAKNRFRVIGDQWIFYGNDGAATNL 1554
RESULT 3
US-10-383-930-40
; Sequence 40, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 40
; LENGTH: 1518
; TYPE: PRT
; ORGANISM: Streptococcus salivarius
US-10-383-930-40
Query Match 46.4%; Score 3818; DB 4; Length 1518;
Best Local Similarity 50.1%; Pred.No.2.4e-193;
Matches 783; Conservative 215; Mismatches 434; Indels 130; Gaps 24;
QY 1 MEKKLHYKLHKVKKHWTVIAVASIGLVSLVG-----AGTVSA---EDKVANDTTAAQTAVG 52
Db 1 MENKIHYKLHKVKKQWVTIAVASVALATVLGGLSVITSSVSADETQDKTVTQNSGTTAS 60
QY 53 VDTGQQAQTNDANTTTDTTADQASANTNODQSGDSNNQOQAQKDTTANTORNQADNS 112
Db 61 LVTSPEATKEADKRTNTKEADVLTPAKETNAVETAT-TTNTQATAEAATATTADVA-VA 118
QY 113 QTDNNQATQOATSPATDGTSTVQRDA--ANTATAADQEGTAPSQEKSAAALSLDNVKL 169
Db 119 AVPNKEAVTTDAPATVTEKAEEQPATVKAEVVNTVEKPAEALKDSEVAALSLKNIKN 178
QY 170 IDGKYVYVQADGSYKKNFAITVNGQMLYFDSDTGALSSSTTSYFSQSGTTLNLDVDFSSHNK 229
Db 179 IDGKYVYVNDGSHKENFALTWNQQLLYFGK-GALTSSTSYSTFTPTTINIVDGFSLNR 237
QY 230 AYDSTAKSFELVNGYLTANSWTRPAGILNGQTWEASNENDLRPVLMSWMPDKDTQVAVY 289
Db 238 AYDSSEASFELIDGYLTADSWYRPAIIKDGVTWQASTAEDFRPLLMWMPNVDVTQVNYL 297
QY 290 NYMNKYL SANETEVTNETSQVDLNKEAQSIQTKIEOKITSDNSTOWLRTAMEAFVAAQPK 349
Db 298 NTWSKVFNL-DAKYSTDKQETLKVAAKDIOIKIEOKIQAEKSTQWLRETTISAFVKTQPK 356
QY 350 WNMSTENFNKG---DHLQGGALLYNTSLTPWANSDYRLNLRPTQDGT--KKYFTEGG 404
Db 357 WNKETENYSKGGEDHLQGGALLYVNDSTPWNDSYRLNLRNTATNTQGTIDKSIDDEQS 416
QY 405 E-----GGVEFLSLNDVNSNPVQAEQLNQLHLYLNNWGDIVMGDKDANFDGVRVDAVDN 460
Db 417 DPNHMGGFDFLLANDVDLSNPVQAEQLNQLHLYLNNWGSIVMGDKDANFDGVRVDAVDN 476
QY 461 NADLLQVNSYFKDNYKVYTDSEANALAHISILEAWSLNDQYNEDTNGTALSIDNSRLT 520
Db 477 DADMLQLYNTYPREYGVNKSSEANALAHISILEAWSLNDHNDKTDGGAALAMENKORLA 536
QY 521 SLAVLTK-----OPGQRIDLSNLISSESVNKEBRAND-----T 551
Db 537 LFLSLAKPIKERTPAVSPLYNNFTNTTQDEKTDWINKDGSKAYNEDGTVKQSTIGKYNE 596
QY 552 AYGDPTPTYSFVRAHSEVQTVIAKIVKEKIDNTSDGYTFTLDQLKDAFKIYNEDMAKNV 611
Db 597 KYGDASGNVYFTRAHDNNVQDIIAEIIKKEINPKSDGFTITDAEMKQAFIYNKMDMLSSD 656

```
QY 612 KYTHYNIIPAAVALLSNMESVPRVYGLYTDGQYMAKSPYDYDAIATMLQRIAYVS 671
Db 657 KKYTLNIPAAVAVMLQNNMETITRVVYGLYTDGHHMETKSPYDYTIIVNLMKSRKIKYS 716
QY 672 GQSEEVH-----KVGNGNQLTSSVRVYGQDLMSADDTQGTDLSTRTSGIATLV 718
Db 717 GGOAQSYWLPTDGKMDNSDELRYRNEVTSYRKGDIIMTANDTEGSKYSRTSGGVTLV 776
QY 719 SNDPNLDLGDGS-LTYNMGRANQAQRPLILGTGQVQSYLKQSD---TNIVKYTDANG 774
Db 777 ANPNKLNLDQSAKLVNEMGKIHANQYRALIVGTADGIKNFTSDADAIAAGYVKETDSNG 836
QY 775 NITFTADDIKGYSTVDMGYLAVVVPVGAQGDVRAADTNQKADGK-SLKTSAAALDSQ 833
Db 837 VLTFGANDIKGYETDFMSGFVAVVVPVGAQSDNQDIRVAPSTEAKKEGELTLKATEAYDSQ 896
QY 834 VIYEGSFNQDF--ANNADATYTKIAENADFKLIGITSFEMAPQYVSATDGSPLDSII 891
Db 897 LIYEGSFNQFTIPDGSQSPVYNNRKIAENVDLFKSGVTSFEMAPOFVSADDTGLDSVI 956
QY 892 QNGYAFSDRYDLAMSNNKNGYKSKDDLALNALKALHANGIOAIADWVPDQIYQLPGEVWTA 951
Db 957 QNGYAFADRYDLAMSNNKNGYKSKEDLRDALKALHAKAGIOAIADWVPDQIYQLPGEVWTA 1016
QY 952 KRTNSYGNPTFDAYINNAYATNTKSSGSDYQAYGGAFLDELKAKYPMFTVMNISTGK 1011
Db 1017 TRTDGAGRKIADAIIDHLSLVANSKSSGKDYOQAYGGEGFLAELKAKYPMFKVMNISTGK 1076
QY 1012 PIDPSTKIQWEAKYFNGTNVLKAGAYVLSDDATCKYTVNENGDFLPASFTGDQNAKT 1071
Db 1077 PIDDSVKLQWKAEBYFNGTNVLERGYYVLSDEATCKYFTVTKBGNFIFLQLTGKEKVI 1136
QY 1072 GFYDGTGMAYYSTGNKAVNSPIYEGGHYYFDDKDHVMVTGYSYKAEDGNDYY-FLPNCI 1130
Db 1137 GFSSDKGITYPTGTSQAQSAFVTENGTYTFDARGHVMVTNSEYSPNGKDVYRFLPNCI 1196
QY 1131 QMRDAIYQDAQNSYVYRGILYKGDNNWYFVDPNNANK-----TVFRYPDANNVMA 1183
Db 1197 MLSNAFYIDANGNTLYNSKGQYKGG--YTRFDVSETDKDGKESKWKVFRYFTNEGVA 1254
QY 1184 IGRNNYGYTYFDENGFOAKGOLLTDDKTHYFDEDNAMAKNFVNGDDWYMDGNG 1243
Db 1255 KGVTVIDGTFQYFGEFGFOAKLVTFKGKTYTFDAHTGNGIKOTWRNINGKWIYFDANG 1314
QY 1244 NAVKGYPVNNQILYFNPTGVQVKGQFITDAQRTSYDYDANSALKSGFFFTPNGSDWY 1303
Db 1315 VAATGAQVINGQKLYFN-EDGSQVKGVVKNADGTYSKYKEGFGELVTFNEFTTDCGNVY 1373
QY 1304 YAEANGVYKGFQVAENQDQWYFDDTQKQAKGAAGVDRDLFYNPDSGVQVKGDFATD 1363
Db 1374 YA-----GANGKTVTGAQVINGQHLFYFNAD-GSQVKGGWKN 1409
QY 1364 ESGNTSFYHDNGDKVVGFFFTGNNAWYADNNGNLVKGQFQIIDGKWHFDEVTQQAQ 1423
Db 1410 ADGTYSKYNASTGERLTNEFFFTGDNWYIYGANGSKSVTGEVKI-----1453
QY 1424 GAALVNGQQLYFDVDSGIOVKGDFVTDGQNTSYDYDNGSKDKVNGFFFTGDNAMWYADG 1483
Db 1454 -----GDDTYFPAKDGKQVKGQTVSAGNGRISYYIGDGGKRAVSTWIELIQPGVYVYFDK 1507
QY 1484 QG 1485
Db 1508 NG 1509
```

RESULT 4

```
US-10-797-821-40
; Sequence 40, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
```

```
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 40
; LENGTH: 1518
; TYPE: PRT
; ORGANISM: Streptococcus salivarius
; US-10-797-821-40
```

Query Match 46.4%; Score 3818; DB 5; Length 1518;

Best Local Similarity 50.1%; Pred. No. 2.4e-193;

Matches 783; Conservative 215; Mismatches 434; Indels 130; Gaps 24;

```
QY 1 MEKLIHYKLHKVKKHWTTIIVASIGLVSLVG-----AGTVSA---EDKVANDTTAAQTAVG 52
Db 1 MENKLIHYKLHKVKKHWTTIIVASVALATVVLGSLVTTSSVSADETDQKTVTQSNSTTAS 60
QY 53 VDTGQDQATNDANTNTDTDADQSANINQDQAGSDQSNQDQAKQDQTANTDRNQADNS 112
Db 61 LVTSPEATKEADKRTNTEADVLTPAKETNAVETAT-TTNTQATAEAATTAITADVA-VA 118
QY 113 QTDNNAQTDQATSPATDGTSGVQRDA--ANTATAADQEGQTAPSEQESKAALSLDNVKL 169
Db 119 AVPNKEAVVTTDAPAVTTEKAEQPAIVKAEVYNTVEVKAPEALDKDSEVEAALSLEIKNI 178
QY 170 IDGKYVYVQADGSKYKNFAITVNGOMLYPDSDTGALSSSTYSFSOGTTNVLVDDFSHNK 229
Db 179 IDGKYVYVNEGDSHKNFAITVNGQLLYFGKD-GALTSSTYSFTPTGTTNIVDGFINNRR 237
QY 230 AYDSTAKSFELVNGYLTANSWYRPAIGILNGOTWEASNENDLRPVLMSWMPKDTQVAYV 289
Db 238 AYDSSEASPELIDGYLTADSWYRPAIIKDGVTWQASTAEDFRPLIMAWPNVDTQVNYL 297
QY 290 NYMKYLSANETEVNTEISOVDLNKEAQSIOTKIEQKITSDNSTOWLRTAMEAFVAAQPK 349
Db 298 NYMSKVFNL-DAKYSSTDKQETLKVAAKDIIQIEQIOAEKSTOWLRETISAFVKTPQ 356
QY 350 WNMSTENFNKG--DHLQGGALLYTNSDLTPWANSDYRLNRTPTQDDGT--KKYFTEGG 404
Db 357 WNKETENYSGGGEDEHLQGGALLYNDSTPTWANSDYRLNRTATNQTGTIDKSLDEQS 416
QY 405 E----GGYEFLLSNVDNSNPVVOAQLNQLHLMNMGDI VMGDKDANFDGVRVDAVDNV 460
Db 417 DENHMGFGPFLANDVLSNPVVOAQLNQLHLMNMGSI VMGDKDANFDGVRVDAVDNV 476
QY 461 NADLLQVYSYFNKDYKVYTDSEANALAHISILEANSLNDNQNEDTNGTALSIDSNSRLT 520
Db 477 DADMLQLYTNYFREYYGVNKSSEANALAHISVLEANSNDNHNKDTDGAALAMENKORLA 536
QY 521 SLAVLTK-----QPQRIDLSNLSSESVNKERAND-----T 551
Db 537 LLFSLAKPIKERTPAVSPLYNNTFTTQDEKTDWLNKDGSKAYNEDGTVKQSTIGKYNE 596
QY 552 AYGDPTIPTYSFVRAHDSYQTVIAKIVKEKIDPNSDGYFTTLDQLDKAFKYNEDMAKNV 611
Db 597 KYGDASGNVYFIRAHNNVQDIIAEIIKKEINPKSDGFTITDAEMKQAFPIYNKMDLSSD 656
QY 612 KYTHYNIIPAAVALLSNMESVPRVYGLYTDGQYMAKSPYDYDAIATMLQRIAYVS 671
```



```
Db 649 SITRLYYGDMYSDGQYMATKSPYYDAIDTLKARIKYAAGQDMKIITVEGDKSHMDW 708
QY 687 ---ILSSVRYGQDLMSADPTQGTDLRSRTSLVTLVNSDNLDLG-GDSLTVMNGRAHANQ 742
Db 709 YTGVLTSVRYGTGANEATD-QGSEATKTQGMVITSNPFLKLNQNDKVIIVNMGAAHKQ 767
QY 743 AYRPLILGTGQVQSYLKDSOT-NIVKYTDANGNLFTTADDIKGYSTVDMSGYLAVWVPV 801
Db 768 EYRPLLLTTKGLTSTDAAKSLYRKNTDNGELVFDASDIQGYLNPQVSGYLAVWVPV 827
QY 802 GAKDQDVRVAADTNQAKDGLSKITSAAALDSOVIYEGFSNFQDFANNDADYTNKKIAENA 861
Db 828 GASDNQDVRVAASNKANATGQVYESSALDSQLIYEGFSNFQDFVTKDSDYTNKKIAQNV 887
QY 862 DFFKGLGITSFEMAPQVYSATDGSFLDSIIQNGYAFSDRYDLAMSKNNKYGSKDDLANAL 921
Db 888 QLFKSGWVTSFEMAPQVYSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSKQDMINAV 947
QY 922 KALHANGIQAIADWPDDQIYQLPGBEVVTAKTNSYGNFTFPAYINNALYATNTKSSGSD 981
Db 948 KALKHSGIQVIAADWPDDQIYNLPGEVVVATRVNDYGEYRKDSEIKNTLYAANTKSNKGD 1007
QY 982 YQAKYGGAFLELDELKAKYPMDFVTNMIISTGKPIDPSTKIKQWEAKYFNGTNVLGKAGYVL 1041
Db 1008 YQAKYGGAFLELDELKAKYPMDFVTNMIISTGKPIDPSTKIKQWEAKYFNGTNVLGKAGYVL 1067
QY 1042 SDAATGKYFTVNEGDFLPASFTGQNAKTGFYDGTGMAYTSTSGNKAIVNFIYEG-GH 1100
Db 1068 KDNASKYFELKGNQTYLPKQMT-NKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGN 1126
QY 1101 YYYFDKGHMVTSYKAEDGNDYFPLPNGIQMRDAIYQDAQNSYYYGRTGILYKGNWY 1160
Db 1127 WYFDFNNGHMVYGLQOL-NGEVQYFLSNGVQLRESPLENADGSKNYFGLHGNRY-SNGYY 1184
QY 1161 PFVDPNNAKTYFRVFDANNWAI GYRNNYGTOTYFEDENGFOAKGOLLTDDKG-THYDE 1219
Db 1185 SF-----DNDKRWYFEDAGWAVGLKTINGNTQYDPDQGYQVKGAWITGSDGKKRYFDD 1239
QY 1220 DNGAMAKNFVN-VGDDWYMDGNGNAVKGOYPVNNQIILFNPETGVQVKGOFITDAQGR 1278
Db 1240 GSGNVAVRFANDKNGDWYLYLNSDGLALVGVTINGKTYFG-QDGKQIKGKIITD-NGK 1297
QY 1279 TSYDANSALKSSGFTPNGSDWYEAENGYYVKGPKVAENQDQWYFDTQTTGQAKGA 1338
Db 1298 LKYFLANSGLARNIPAT-----DSQNNWYFEG-SDGVAVTGS 1334
QY 1339 AKVDGRDLVFPDPSGVQVKGDFATDESQNTSFVHGDNQKVVGGRPTTCNNAWYADNNG 1398
Db 1335 QTIAGKKLYFASD-GKQVKGSEFT-YNGKVHYTHADSGELQVNRFPADKQGNWYLYDSNG 1392
QY 1399 NLVKQFQEIDGKWHYFDEVTGQQAALVNGQQLYFVDSDIGIQVKGDFVTD 1450
Db 1393 EALTGSQRI-----NDQRVFF-TREGQVKQGDVAYD 1422
```

RESULT 6

```
US-10-383-930-36
; Sequence 36, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 36
```

```
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-36

Query Match 43.9%; Score 3615; DB 4; Length 1430;
Best Local Similarity 49.9%; Pred. No. 1.2e-182; Indels 112; Gaps 32;
Matches 745; Conservative 217; Mismatches 418;

QY 1 MEKKLHYKHLHKVKGHWVTIAVASIGLVSLVAGAGTVSAEDKLVANDTTAQATVGVDTCQ--- 57
Db 1 METKERYKMHKVKHVTIVAVAS-GLITL---GTTTLGSSVSAETEQQTSKVKVTKQKSD 56
QY 58 DOATNTDANTVNTDTDTADQASANTNODQAGSQSNNOQAKQDT-----ANTOR 106
Db 57 DKAASESQTDAPKT---KQAQTEQTOAQS-QANVADTSTSTITKETPSQNTITQANSDD 111
QY 107 NOADNSQTDNNQATDOATSPATDGTSVQRRDAAVATA-ADQEGOTAPSEQEKSAAL-SL 164
Db 112 KTVNTTKSEBAQTSBEERTKQSEEAQTTASSQALTQAKAELTKQRTAAQENKNPVDLAAI 171
QY 165 DNVKLDIGKYYVQADGSGYKKNFAITVNGQMLYFDSDTGALSSSTSYFSQGTITNLVDDF 224
Db 172 PNVKQIDGKYYIGSDGQPKKNFALTANVKVLYFDKNTGALTDTSTQYQFKQGLTKLNNDY 231
QY 225 SHNKAYDSTAKSFELVNGYLTANSWYRPAGILRNQGTWEASNENDLRPLVMSWPPDKOT 284
Db 232 TPNQIVNFENTSLETDIDNYVTADSWYRPPKQILKNGKTWTASSESDLRPLMSWPPDKOT 291
QY 285 OVAYVYNNKYLSANETEVTNETSQVDLNKEAQSQTKEOKITSDNSQWLTARMEAFV 344
Db 292 QIAYLNYNNQGLGTGENYTDSSQESLNLAATQVQVKIETKISQTOQTQWLRDIINSFV 351
QY 345 AAQPKWNMNSTE---NFNKGDLHQQGALLYTNSDLTPWANSYDRLNLRPTQODGTGKYFT 401
Db 352 KTOPNWNQTESDTSAGEKDLHQQGALLYSNSDKTAYANSYDRLNLRPTSTQGTGPKYFE 411
QY 402 EGEGGYEFLNSDNDVNSNPVQABQLNQLHYLMNWGDI VMGDKDANFQGVRVADVNVN 461
Db 412 DNSSGYDFLLANDIDNSNPVQABQLNWLHYLMNYGSI VANDPEANFQGVRVADVNVN 471
QY 462 ADLLQVYNNYKDYKVTDDSEANLAHLSILEAWSINDNQYNETNGTALSIDNSRLTS 521
Db 472 ADLLQIADSYLKAHYGVDKSEKNAHLSILEAWSINDNQYNETNGTALSIDNSRLTS 531
QY 522 LAVLTQPCQQRIDLNS-----LISESVKNERANDTAYGDTIPTYSFVRAHDSEVQ 571
Db 532 LYALTR-PLEK-DASNKNEIRSGLEPVTNSLN-NRSABGKNSEMANIYFIRAHDSVQ 588
QY 572 TVIAKIVKEIDTNSDGYFTFLDQKDAFKIYNEDMAKVNKYTYTHYNIIPAAVALLLSNME 631
Db 589 TVIAKIIKAQINPKTDGLTFTLDELKQAFKIYNEDMRQAKKYYTOSNIPTAYALMLSNKD 648
QY 632 SVPRVYVYGDLYTDDGQYMAKKSPPYDAITATMLQGRIVAVYVSGGSEBHVKNVGNQ----- 686
Db 649 SITRLYYGDMYSDGQYMATKSPYYDAIDTLKARIKYAAGQDMKIITVEGDKSHMDW 708
QY 687 ---ILSSVRYGQDLMSADPTQGTDLRSRTSLVTLVNSDNLDLG-GDSLTVMNGRAHANQ 742
Db 709 YTGVLTSVRYGTGANEATD-QGSEATKTQGMVITSNPFLKLNQNDKVIIVNMGAAHKQ 767
QY 743 AYRPLILGTGQVQSYLKDSOT-NIVKYTDANGNLFTTADDIKGYSTVDMSGYLAVWVPV 801
Db 768 EYRPLLLTTKGLTSTDAAKSLYRKNTDNGELVFDASDIQGYLNPQVSGYLAVWVPV 827
QY 802 GAKDQDVRVAADTNQAKDGLSKITSAAALDSOVIYEGFSNFQDFANNDADYTNKKIAENA 861
Db 828 GASDNQDVRVAASNKANATGQVYESSALDSQLIYEGFSNFQDFVTKDSDYTNKKIAQNV 887
QY 862 DFFKGLGITSFEMAPQVYSATDGSFLDSIIQNGYAFSDRYDLAMSKNNKYGSKDDLANAL 921
Db 888 QLFKSGWVTSFEMAPQVYSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSKQDMINAV 947
```

QY 922 KALHANGIOAIADWPDQIIVOLPGEVVTAKRTNSYGNPTFDAYINNALYATNTKSGSD 981
DB 948 KALHKSQIOAIADWPDQIIVLPGKEVVTATRVNDYGEYRKDSSEIKNTLYAANTKSGKD 1007
QY 982 YQAOYGAFDELKAKYKYPDMFTVMNISTGKPIDPSTKI KOWEAKYFNGTVNLGKAGYVL 1041
DB 1008 YQAOYGAFSELAAKYPSIFNRTQISNGKKIDPSEKITA KAWKIFNGTINILGRGVYVL 1067
QY 1042 SDDATGYFTVNEGDELPAFTGDQNAKTGFYVDGTGMAYYSTSGNKAVNSFIYEG-GH 1100
DB 1068 KDNASDKYFELKGNQTYLPKQMT-NKEASTGFVNDGNGMTFYSTSGYQA KNSFVQDAKGN 1126
QY 1101 YYYFDKCHMVTGSYKAEDGNDYFYPNGIQMRDAIYQDAGNSYYGRTGILYKGDWY 1160
DB 1127 WYYFDNNGHMYGLQQL-NGEQVYFLSNGVQLRESEFLENADGSKNVEFGLGNRY-SNGYY 1184
QY 1161 PFVDPNNANKTVPRYPDANNMAIGVRNMYGOTYFDPDENGFOAQGLTDDKG-THYFDE 1219
DB 1185 SF-----DNDSKRYFDPASGVMAVGLKTINGNTQYFDQDGIQVKGAWITGSDGKRYFDD 1239
QY 1220 DNGAMAKNFVN-VGDDWYMDGNGNAVKGOYFVNNQILYFNPETGVQVKGQFITDAQGR 1278
DB 1240 GSGNMAVNRFPANDKNGDWYLLNSDGIALVGQTINGKTYFYG-QDGKQIKGKIITD-NGK 1297
QY 1279 TSYDANSALKSSGFPTPNDSWYYAENGYYVYKFKQVAENQDQWYFDPQTGKQAKGA 1338
DB 1298 LKTYFLANSGBELARNIFAT-----DSQNNWYFYG-SDGVAVTGS 1334
QY 1339 AKVDGRDLYPNDSGVQVKGDFATDSGNTSFYHGONGDKVGGPFTTGNNANYADNG 1398
DB 1335 OTTAGKKLFPASD-GKQVKSFTV-YNGKHYHADSGBELQVNRFEADKDGNNWYLDNSG 1392
QY 1399 NLVKGQIEIDGWYHFDVETGQAKGAALVNGQQLYFVDVDSGIQVKGDFVTD 1450
DB 1393 EALTGSQRI-----NDQRVFF-FREGQVKGDVAYD 1422

RESULT 7

US-10-797-821-36
; Sequence 36, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 36
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-36

Query Match 43.9%; Score 3615; DB 5; Length 1430;
Best Local Similarity 49.9%; Pred. No. 1.2e-182;
Matches 745; Conserved 217; Mismatches 418; Indels 112; Gaps 32;

QY 1 MBKCLHYLHKVKGHWVTIAVASIGLVSLVGAGTVSAEDKAVANDTTAQTAVGVDTGQ---57

DB 1 METRRYKHKVKGHWVTIAVAS-GLITL---GTTILGSSVSAETEQTSDKVVTKQSED 56
QY 58 DQATTNDANTNTTDTADQASANTNOQAGSDOSNNQDAKQDT-----ANTDR 106
DB 57 DKAASSQTDAPKT---KQAEQTQAOQS-QANVADTSTSTKETPSSNITTOQANSDD 111
QY 107 NQADNSQTDNQATQATSPATDGTISVQRREDAANVATA-ADQEQQTAPSQEAKSAAL-SL 164
DB 112 KTVNTKSEBAQTSSEERTKQSEBAQTASSQALTKQAKAELTKQRTAAQENKPNVDLAAI 171
QY 165 DNVLKIDGKYIYQADGSKYKGNFAITVNGOMLYFSDSTGALSSSTSTSYFSQGTNNLVDDF 224
DB 172 PNYKQIDGKYIYIGSDQPKKNPALVNNKVLFDKNTGALTDTSTQYFQGLTKLNDY 231
QY 225 SSHKAYDSTAKSFELVNGVLTANSVWRPAGIILRNQGTWEASNENDLRPVLMSWPPDKOT 284
DB 232 TPNQIIVNTEFNTSLETIDNYVTADSWYRPNKIDILNKGKTTWASSESDURLPLMSWPPDKOT 291
QY 285 QVAYVNMNKLNSANETEVNTNETSQVDLNKEAQSIQTKI BQKITSDNSTQWLRTAMBAFV 344
DB 292 QIAYLNMNQOGLGTGENYTADSSQESLNLAAQTVQVKIETKISQTOQWLRIINSFV 351
QY 345 AAOQKWNMSTE---NFNKGDLQGGALLYTNSDLTPWANSDYRLLRNTPPTQOQGTKKYFT 401
DB 352 KTOPNNSQTESDTSAGEXDHLQGGALLYSNSDKTAYANSYRLLRNTPPTSTQTKPKYFE 411
QY 402 EGEGEGYEFLSNDVNSNPVQAEOLNOLHYLMWNGDIVMGDKDANFQDVRDADVNNV 461
DB 412 DNSGGYDFELLANDIINDSNPVPVQAEOLNMLHYLMNYGSIIVANDPEANFQDVRDADVNNV 471
QY 462 ADLLQVYVYFNQYKVTDSANALAHISILEAWSLNDNOYNDTNGTALSINSSRLTS 521
DB 472 ADLLQIADSLYLAHYGVDSKNAIINHLISILEAWSNDNDPQYNKDTKGAQLPIDNKLRLSL 531
QY 522 LAVLTQPGQRIDLSN-----LISESVNKERANDTAYGDTTPTTYSFVRAHSEVQ 571
DB 532 LYALTR-PLEK-DASNKNBIRSGLEPVITNSLN-NRSAEGKNSERMANVYIFIRAHSEVQ 588
QY 572 TVIAKIVKEIDTNSDGYTFTLDOLKDAKINEDMAKNVKTTHYNI PAAYALLSNME 631
DB 589 TVIAKIIKAIQINPKTDGLTFTLDELQAFKIYNEDMRQCKKYTQSNIPAYALMLSNKD 648
QY 632 SVPRVYIGLYTDDGOMAKKSPYDAIATMLQGRIAIYVSGGSEEVHKVNGNQ-----686
DB 649 SITRLYIGDMYSDGQYMATKSPYDAIDTLKARIKAAAGQGMKITTYEGDKSHMDWD 708
QY 687 ---ILSSVRYGQDLMSADDTQGTDLRSRTSLVTLVSNPNLDLG-GDSLTVNMGRAHANQ 742
DB 709 YTGVLTSVRYGTGANEATD-QGSEATKTQGMVITSNPSPSLKLNQNDKVIIVNMGAAHKQ 767
QY 743 AYRPLILGTGDKGVQSVYLKDSDT-NIVKYTDANGNLFTTADDIKGYSTVDMSGYLAVVPV 801
DB 768 EYRPLLLTTKDLTSTSYSDAAAKSLYRKNTNDKGELVFDASDIQGLYNLPQVSGYLAVVPV 827
QY 802 GAKGQDVRVAADTNQKADGKSLKTSALDSOVIYEGFSNFQDFANNADYTNKKIAENA 861
DB 828 GASDNDQVRVAASNKANATGVYESSALDSQIYIEGFSNFQDFTVKDSQDYTNKKIAQNV 887
QY 862 DFFKGLGITSFEMAPQYVSATDGSFLDSIIQNGYAFSDRYDLAMSNNKNGSKDDLANAL 921
DB 888 QLFKSGVTSFEMAPQYVSSDGSFLDSIIQNGYAFEDRYDLAMSNNKNGSKQDDMINAV 947
QY 922 KALHANGIOAIADWPDQIYQLPGBEVVTAKRTNSYGNPTFDAYINNALYATNTKSGSD 981
DB 948 KALHKSQIOAIADWPDQIYNLPGEVVTATRVNDYGEYRKDSSEIKNTLYAANTKSGKD 1007
QY 982 YQAOYGGAFLDELKAKYKYPDMFTVMNISTGKPIDPSTKI KOWEAKYFNGTVNLGKAGYVL 1041
DB 1008 YQAOYGGAFLSELAAKYPSIFNRTQISNGKKIDPSEKITA KAWKIFNGTINILGRGVYVL 1067
QY 1042 SDDATGYFTVNEGDELPAFTGDQNAKTGFYVDGTGMAYYSTSGNKAVNSFIYEG-GH 1100

Db 1068 KDNASDKYFELKGNQTYLPKQMT-NKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGN 1126
QY 1101 YYYFDKDGHWVTGSKYKAEDGNDYYFLPNGIQMRDAILYQDAQNSYYTGRGILYKGDWY 1160
Db 1127 WYFYDNNGHMVYGLQOL-NGEVQYFLSNGVQLRESFLENADGSKNYFGLGNRY-SNGYY 1184
QY 1161 PFVDPNNAKTVPRYFDANNVMAIGYRNMVGTQYYFDENGFOAKGOLLTDDKG-THYFDE 1219
Db 1185 SF-----DNDKRWYFDASGWWAVGLKTINGNTQYFDQGYQYKGAWITGSDGKKRYFDD 1239
QY 1220 DNGAMAKNFVN-VGDDWTYMDGNGNAVKQYQYVNNQIILYFNPETGVQVKQGITDAQGR 1278
Db 1240 GSGMAVNFANDKNGDWYLLNSDGLALGVQVITNGKTYFYG-QDCKQIKGLIITD-NGK 1297
QY 1279 TSYDANSALKSSGFFTPNGSDWYEAENGYYVYKGFQVAENQDQWYFDDTGTGQAKGA 1338
Db 1298 LKYFLANSGLARNIPAT-----DSQNNWYFYG-SDGVAVTGS 1334
QY 1339 AKVDGRDLVFNPSGVQVKGFATDESNTSFYHGDNGDKVVGFFTTGNNAWYYADNG 1398
Db 1335 QTITAGKGLYFASD-GKQVKGSTV-YNGKHYHADSGELQVNRFEADKDGWYLLDSNG 1392
QY 1399 NLVKGFQEIDGKWHFDEVTGQAKGAALVNGQOLYFVDVSDGIVQKGFVTD 1450
Db 1393 EALTGSQRI-----NDQVFF-FREGKQVKGDVAYD 1422

RESULT 8
US-10-383-930-37
; Sequence 37, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 37
; LENGTH: 1590
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-383-930-37

Query Match 42.5%; Score 3502.5; DB 4; Length 1590;
Best Local Similarity 45.4%; Pred. No. 1.3e-176;
Matches 749; Conservative 252; Mismatches 482; Indels 167; Gaps 40;
QY 1 MEKLLYKHLKVKKHWVTAVASIG-LVSLVAGTVSAEDKVANDTTAQATVGVDTGQDQ 59
Db 1 MEKNVRFKHKVKRWVTLVSASATWLASALGASVASADTTASDDSNQAVV---TG-DQ 56
QY 60 ATTND--NTNTTDTDTAQASNTQDQAGSDQSNQDQAKQDQANTDRNQADNSQTDNN 117
Db 57 TTNQATDQTSIAATATSEGSASTD---AATDQASAEQTQGTASTD--TAAQTITNAN 111
QY 118 QATDQATSPATDG-TSVQRDAAANVATAAQEGQTAPEBOEKAALSLONVKLIDGKYYY 176
Db 112 EAKWVPTENENQGTDEMLAEAKNVATA---ESDSIPSDLAK-----MSNVKQVDGKYYY 163
QY 177 VQADGSKYKNFAITVNGQMLYFSDTGALSSSTYSYFSQGTNNLVDD---FSSHKNAYDS 233
Db 164 YQDGGNKKKNFAVSGVDKIIYFD-ETGAYKQDTSKVDADKSSSAVSQATIFAANNRAYST 222
QY 234 TAKSFELVNGYLTANSWYRPAIGLRNGQTWEASNENDLRPLMSWMPDKDQTVAYVYNN 293
Db 223 SAKNFEAVNYLTADSWYRPKSLKDGKTWESGKDDFRPLLMAWPPDITETKENYVYNN 282

QY 294 KYLSANETEVTNETSQVDLANKEAQSIQTKIEOKITSDNSTOWLRTAMEAFVAAQPKWMS 353
Db 283 KVVGDGKT-YTAETSQADLTAAAEVLQARIEQIKISENNTKWLREIAISAFVTKPQOMNE 341
QY 354 TENFNKGDLHQQGALLYTN-SDLTPWANSDYELLNARTPTQODGTK-----KYFTEGEGGY 408
Db 342 SEK-PYDDHLQALLFNDQTLTPTQSNRYLLNARTPTNQTSGLDSRSTYFNPNDPLGGY 400
QY 409 EFLANDVDSNPVQAEOLNQLHYLMNMGDIIMGDKDANFDGVRVDAVDNWNADLLQVY 468
Db 401 DELLANDVDSNPVQAEOLNWLHYLLNFGSIYANDADANFDSIRVDAVDNVDADLLQIS 460
QY 469 SNYFKDNYKVTSEANALAHISILEAWSLANDQYNEDTNGTALSIDNSSRLTSLAVLTQ 528
Db 461 SDYLLKAAAGIDKKNANNHVSIVEAWSNDPTFYLHDDGDNLMNMNDKFLSLMLSLAKP 520
QY 529 PQCRIDLSNLISSESYNKERANDTAYGDTIPTYSFVRAHDSVQTVIAKTVKEKIDTNSDG 588
Db 521 LDKRSGLNPLIHNSL-VDEVDREVEVTPSYSPARAHDSVQDIIRDIIKAEINFSFG 579
QY 589 YFTTLDQLKAPKIYNEDMAKVNKYTHYNIIPAAVALLSNMESVPRVYVYGLYTDGQY 648
Db 580 YSFTQEEIEQAFKIYNEDLKKTKYTHYNVPLSYLLTNKGSIPRVYVYGMFTDDGQY 639
QY 649 MAKKSPPYDAIATMLQGRITAYUSGQSEBHVKNVGNNOILSSVRYQODLMSADDTQGTDL 708
Db 640 MANKTVNYDAIESLLKARKMYSGGQAMQNYQI-NGGEILTSTRYRGKALKQSD-KGDAT 697
QY 709 SRTSGLVTLVSNDPNLDLGGDSLTVNMGRAHANQAYRPLILGTQKQVQSYLKDSDTN--- 765
Db 698 TRISGVVVGNGQPNFSLDGKVVALNMGAAHANQAYRALMVSTKQGVATYATDADASKAG 757
QY 766 IVKYTEANGNLFTTADDKYGVSTVDMSGYLAVMVPVGAQKQDQVRVAADTNQKADGSKLK 825
Db 758 LVKRTDENGLYFLNDDLKGVANPQVSGFLQVWPVGAADDQDIRVAASDTASTDGKSLH 817
QY 826 TSAALDSQVIYEGPSNFQDFANNADADYTNKKTAEADPFKLGITSFENAPQVYSATDGS 885
Db 818 QDAAMDSTRVMEGFSNFQSFATKEBEYTNVTVANNVDFKVSWGITDFEMAPQVVSSTDG 877
QY 886 FLDSIIQNGYAFSDRYDLAMSNNKYGSKDDLANALKALHANGIOAIAADVPDQIYQLPG 945
Db 878 FLDSVIQNGYAFTRDYDLGMSKANKYGTADQLVKAIKALHAKGLKVMADVPDQMTFFK 937
QY 946 EEWVTAKRTNSYGNPTFDAYINNLYATNTKSSGSDYQAYGGAFLDELKAKYPMFTVN 1005
Db 938 QEVVTVTRTDKFGKPIAGSQINHSLVYVTDTKSGSDDYQAKYGAFLDELKAYPELFTKK 997
QY 1006 MISTGKPIDPSTKIQWEAKYFNGTNNLKGAGYVLSDDATGKYFTVNGENDFLPASFTG 1065
Db 998 QISTGQALDPSVKIKQWSAKYFNGSNILGRGADYVLSQVSNKYFNVASDITLFLPSLLG 1057
QY 1066 DONAKTGFYDGTGMAYYST-SGNKAVNSFIYEGGHYVYFDKDGHWVTGSKYKAEDGNDY 1124
Db 1058 -KVESGIRYDQKGIYNSSATGQDVQKASFIIEAGNLYYFGDKDGYNVTGA-QTINGANF 1115
QY 1125 FLPNGIOMERDAIYODAQGNSYYTGRGILYK-----GDNWYFVDPNNANKTVFYPDA 1178
Db 1116 FLENGTALRNTIYTDQAQGNSHYVANDGKRYENGYYQFGNDW-----RYFKD 1161
QY 1179 NNVMAIGYRNMVGTQYYFDENGFOAKGOLL-TDDKTHYFDEBNGAMAKNFV-NVGDWW 1236
Db 1162 GN-MAVGLTVDGNGVQYFDKQGVQAKDKIIVTRDQKRVYFDQHNNGNAANTFIADKTGH 1220
QY 1237 YMGDGNAGVAKQYVNNQILYFNPETGVQVQGITDAQGRTSYYDANSGLAKSSGFFT 1296
Db 1221 YYLKDGVAVTAQTVGKQKLYFE-ANGQVKGDFVTSDEGKLYFYFDVDSGDWMTDTTIE 1279
QY 1297 PNGSDWYEAENGYYVYKGFQVAENQDQWYFDDTGTGQAKGAQVDRGLRFLNPNDSGVQV 1356
Db 1280 DRAGNWFYLGK-----DGAAVTAQITRGQKLYFKA-NGQOV 1315

[illegible]

RESULT 9

```

US-10-797-821-37
; Sequence 37, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 37
; LENGTH: 1590
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-797-821-37

```

Query Match	42.5%;	Score	3502.5;	DB	5;	Length	1590;
Best Local Similarity	45.4%;	Pred. No.	1.3e-176;				
Matches	749;	Conservative	252;	Mismatches	482;	Indels	167;
						Gaps	40;

QY	1	MEKKHLVHLVKVKKHWVTIAVASIG-LVSLVGGAGTYSAECKVANDTTTAAQTATVGVDTCODQ	59
DB	1	MEKNVRFKHVKVKKRWVTLVSASATMLASALGASVASADTDTASDDSNQAV---TG-DQ	56
QY	60	ATTNDA--NTNTTDDTTADOSANTNQDAGSDOSNQDOAKQDTANTDRNQANQSOTDNN	117
DB	57	TTNNQATDQTSIAATATSEQASD---AATDQASAAEQTGGTASTD---TAAQTTTNAN	111
QY	118	QATDQATSPATDG--TSVQRDAAVATAADQEGTAPSEQEKSAAALSLDNVKVLIDGKYYY	176
DB	112	EAKWVPTENENQGTDEMLAEAKNVATA---ESDSIPSDLAK-----MSNVKQVDGKYYY	163
QY	177	VQADGSYKKNFATVNGOMLYFPDSDTGALSSSTYSFSQGTTHLVD-----FSSHNKAYDS	233
DB	164	YDQDGNVKKGNKFAVSVGDKIYYFD--ETGAYKDTSKVDADKSSSSAVSQNATIFAANNRAYST	232
QY	234	TAKSFELVNGYLTVANSWYRPAGILRRGTQWEASNENDLRPLVMWPPDKDTQVAVVNYMN	293

Db	223	S	AKNFEAVDNYLTADSWYRPKSLKDGKWTESGKDDFRPLLMAWWPDTEFKRYVNYMN	282
Qy	294	K	YLSANETEVTNETSQVDLKAQSIQTKIEQKITSDNSTQWLRTAMEAFVAAQPKWNMS	353
Db	283	K	VVGIDKT-YTAETSQADLTAABELVQARLEQKITSENNYKWLREALSAFVKTOPQWNGE	341
Qy	354	T	ENPNKGDLHGQGGALLVTN-SDLTPWANSYRLLNRPTQODGPK-----KYFTEGGEGGY	408
Db	342	S	EK-PYDDHLQNGALLFDNQTDLTPPQTSNYRLLNRPTQTSGLDSRFTYVNPNDPLGGY	400
Qy	409	E	FLLSNDVDNSNPVQAEQLNQLHYLNMWGDIVMGDKDANFDGVRVDVAVNVDNADLLQVY	468
Db	401	D	FLLANVDVNSNPVQAEQLNQLHYLNFSGIYVANDADANFDSIRVDVAVNVDNADLLQIS	460
Qy	469	S	NYFPKONYKVTDSANALAHISILEAWSLNDNQYNEDTNGTALSINDSSRLTSLAVLTQK	528
Db	461	S	DYLYKAAAGYDKNNKNANNHVSIVEAWSNDTPLYLHDDGDNLMNMNDKFRLSMLWSLAKP	520
Qy	529	P	QGORIDLNLISBSVKNKERANDTAYGDTITPYGFPVRAHDSVQTVIAKIVKEKIDTNSDG	588
Db	521	L	DKRSGLNPLIHNLSL-VDRREVDRVETVPSYFARAHDSVEQDIIRDIIKAEINPNPSFG	579
Qy	589	Y	TFTLDQLKDAFKTYNEDMAKVNTYTHYNI PAAYALLLNMESVPRVYVYGDLYTDDGOY	648
Db	580	Y	SFTQEEIEQAFKIYNEDLKKTKYTHYNVPUSYTLULLNKGSI PRVYVYGDMPFTDDGOY	639
Qy	649	M	AKKSPYYDAIAITWLOGRIAYVSGGSEEVHKVGNQNOILSSVRYQGDLMSADDTQGTDL	708
Db	640	M	ANKTVNYDAIESLLKARKMYVSGGQAWQNYQI-NGEILITSRYVKGALKQSD-KGDAT	697
Qy	709	S	RTSGLVTVLSNDPNLDLGDSTITVNMGRAHANQAYRPLILGTGKGQVQSYLKQSDTN---	765
Db	698	T	RTSGVGVWVGQNFSLDCKVVALNMGAAHQAQVEYRALMVSTKQGVATYATDADASKAG	757
Qy	766	I	VKYTDDANGNLTTADDIKGYSTVDMSGYLAWVPVCGAKQDQVRVAADTNQKADGSLK	825
Db	758	L	VKRTDENGLYLFLNDDLKGVANPQVSGFLQVWPVPVGAADDQDIRVAASDTASTDGKSLH	817
Qy	826	T	SAALDSOVIYEGFSNFQDFANNADVTNKKIAENADFFKKLGITSPEMAPOVYSATDGS	885
Db	818	Q	DAAMGRVWPEGFSNFSQSFATKEEYTNVVIANNVDKFSWGLTDEMAPOVYSSTDGG	877
Qy	886	F	FLDSIIQNGYAFDRYDLAMSNNKYGSKDDLANALKALHANGIOAIADWVPDQIYOLPG	945
Db	878	F	FLDSVIQNGYAFDRYDLGMSKANKYGTADQLVKAIKALHAKGLKVMADWVPDQMYTFPK	937
Qy	946	E	VVUTAKRTNSYGNPFTDAYINNALLYATNTKSSGSDYQAQYGGAFIDELKAKYPPDMPTVN	1005
Db	938	Q	EVVTVTRTDKFGPIAGSQIHNLSLYTDTDKSSGDDYQAKYGGAFIDELKEXYPELFTPK	997
Qy	1006	M	ISTGPIDPSTKIKOWEAKYFNGTIVLHGKGYVLSDDATGKYFTVWNGDFLPASFTG	1065
Db	998	Q	ISTGQAIDPSVKIKQWSAKYFNGSILNLRGADYVLSDQVSNKYFNVASDTLFLPSLLG	1057
Qy	1066	D	QNAKTGFYDGTGMAIYST-SGNKAVNSFIYEGGHYFYFDKGDHMTGSYKAEADGNDYY	1124
Db	1058	-	KVVEGIRYDGGYIYNSSATGDQVKASPI TEAGNLYYFGKQGVMTGA-QTINGANYF	1115
Qy	1125	F	PLNGIQMRDAIYQDAQGNSSYYYGRTGILYK-----GDNWYFPVDPNNANKTVFRYFDA	1178
Db	1116	F	LENGTALRNTIYTDAGNSHYVANDGKRYENGYYQQFGNDW-----RYFKD	1161
Qy	1179	N	NYWATGYRMYQTYVYFDENGFOAKGOLL-TDDKTHYFDENDGAMAKNKFV-NVGDDW	1236
Db	1162	GN	-MAYGLTVDGNVOYFDKQGVQAKDKIIVTRDGVKRVYFPQDNGNNAATNFTADKTGHW	1220
Qy	1237	Y	YMDGNNGNAVKGQYVNNOLLYFNBPETGVQVKGOFITDAQRTSYSDANSALKSGSFFT	1296
Db	1221	Y	YLGKGVAVTGAQTVGKQLYFE-ANGQVKGDFTVSEGLKYFYFDVDSGDMMWTDTFIE	1279
Qy	1297	P	NGSDWYAEANGYVYKGFQVABENQDQWYFYDQTTGQAKGAARVQDGRDLYFNPDPSGVQV	1356

```
Db 1280 DRAGNWFYLGK-----DGAAVTGAQIRGQKLYFKA-NGQOV 1315
QY 1357 KGFAPDESGNTSFYHGDNGDKVVGFFTTGNNAWYADNG-----NLVKG-----1403
Db 1316 KGDIVRGTDKIRYDAKSGEQVFNKVAADGKTVIIGNDGVADVPSVVGQTFFKXASG 1375
QY 1404 ---FQEIDGK-----WY---HFDEV---TGQAKAALVNGOOLYFDVDSGIVKQKDFV 1448
Db 1376 AURFYNLKGOLVTGSGWYETANHDWYIIOGKALTGEQTINGOHLFYKED-GHQVKGOLV 1434
QY 1449 TDQOGNTSYVDNSGDKVNGFFTTGDNWYI-----1480
Db 1435 TGTDGKRVYYDANGSGDAFNKSVTVNGKTYFNGNDGTAQTAGNPKGQTFKDGSDIRPFYSM 1494
QY 1481 -----ADGO-----GNLAKGRKSIDNQDLYFDPATGKQVKGOLV-SIDGRN 1520
Db 1495 EGQLVTGSGWYENAOQOWLYVKNKGVLTLGLQTVGSQRVYFD-ENGIQAKGKAVRTSDGKI 1553
QY 1521 YFPDSGSGMAKNRFVIRIGDOWLYFGNDGA 1550
Db 1554 RYFDENSGSGMITNQWKFVYGYQYYFNGDGA 1583

RESULT 10
US-09-740-274-2
; Sequence 2, Application US/09740274
; Patent No. US2002031826A1
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740, 274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-740-274-2

Query Match 41.7%; Score 3434.5; DB 3; Length 1475;
Best Local Similarity 46.6%; Pred. No. 4.6e-173;
Matches 731; Conservative 248; Mismatches 458; Indels 131; Gaps 39;

QY 1 MEKKLHYKLHKVKKHWVTAVAS-IGLVSLVGAGTVSAEDKVANDTTAQTAVGVDGQDQ 59
Db 1 MDKKVRYKURKKRWVTVSVAWTLTTLGGLVKAD---SNESKSQ-----46
QY 60 ATTNDANTNTDPTDADQSAN-----TNODQAGSDQSNQDQAKQDTAN-----103
Db 47 -ISNDSNTSVV---TANESNVITEATSKQEAASSQTNHTVTTSSSSTSVNPKVEVSNP 102
QY 104 -TDNRQADNSQTDNNQATQATSPATDGTGTVQRDDAANVATAADQSGQTAPSEQKSAAL 162
Db 103 YTVGETASNGEKLQNTT-----TVDKTS-----EAAANNISKOTTEADTDVIDDSNAANL 153
QY 163 ----SLDNVKLIDGKYVYVQADGSYKKNFAITVNGOMLYFDSDTGALSSTSTYSFSQG-T 217
```


Db 1365 -----TGARTINGQHLYFRA-NGVQVKGFEVTDHGRHSIYDNGSGDQ 1406
QY 1466 KVGFFFTGDNAMYADGOGNLAGRKSIDNQDLYFDPATGKQVKGQLYSID-GRNYIYED 1524
Db 1407 IRNFRVNAQGWFFYFDNNGYAVTGARTINGQHLYF-RANGVQVKGFEVTDYGRISYYD 1465
QY 1525 SGSGNMAK 1532
Db 1466 ANSGERV 1473
RESULT 12
US-10-797-821-34
; Sequence 34, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 34
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-34

Query Match 41.7%; Score 3434.5; DB 5; Length 1475;
Best Local Similarity 46.6%; Pred. No. 4.6e-173;
Matches 731; Conservative 248; Mismatches 458; Indels 131; Gaps 39;
QY 1 MEKLLHYKLHKVKKHVVITAVAS-IGLVSLVGAGTVSAEDKVDANDTTAQATVGVDITGQDQ 59
Db 1 MDKKVRYKLRKKVKKRWTVSVASAVMTLTTLGGGLVKAD---SNESKSO----- 46
QY 60 ATTNDANTNTDITDADQSAN-----TNQDQAGSDQSNQDQAKQDTAN----- 103
Db 47 -ISNDSNTSVV---TANESNVITEATSKQEAASSQTNHVTSSSSTSVNPKVVSNP 102
QY 104 -TDRNQADSQTDNNQATQATSPATDGTGSVQRDDAANVATAADQEGQATPSEQEKAAL 162
Db 103 YTVGETASNGEKLQNTT-----TVDKTS---EAAANISQTTTEADTDVIDDSNAANL 153
QY 163 ----SLDNVKLIDGKYVYVQADGSYKKNFAITVNGOMLYFSDTQALSTSTVSFSQG-T 217
Db 154 QILEKLPNVEIDGKYVYVYDNNQKRVNTFTLIADGKILHFD-ETGAYTDTSDTVNKOIV 212
QY 218 TNLVDDFSHNKAYDSTAKSFELVNGYLTANSWYRPAGILRNGQTWEASNENDLRPVLMS 277
Db 213 TTRSNLYKKYNQVYDRSAQSFHVDHLYTAESWYRPKYILKDGKWTQTSTEXDQFRLMT 272
QY 278 WPPDKDTQVAYVNNMKYLSANETEVNTNETSQVDLNKEAQSQTQKIEQKITSNQTQWLR 337
Db 273 WPPDQETQRYVNNMAQLGINKT-YDDTSNQLNIAAAATQAKTEAKITTLKNTDWLR 331
QY 338 TAMEAFVAAQPKWNMSTEN-FNKGDLHQSGGALLYTN-SLTPWANSDYRLNARTPTQDQ 395

Db 332 QTISAFVKTQSAWNSDSEKPPD--DHLQNGAVLYDNEGKLT.PYANSNYRILNRTPTNQTG 389
QY 336 TK--KYFTEGEGGGEFFLLSNDVNSNPVQABQLNQLHYLWNGDILVNGDKDADPDGVR 453
Db 390 KKDPRTADNTTGGYEFLLANDVNSNPVQABQLNQLHYLWNGDILVNGDKDADPDGVR 449
QY 454 VDADVNVADLLOVYSNYFKDNYKYVTDSEANALAHISILEAWSLNDNQYEDTNGTALGI 513
Db 450 VDADVNVADLLOVYSNYFKDNYKYVTDSEANALAHISILEAWSLNDNQYEDTNGTALGI 509
QY 514 DNSSRLLTSIAVLTKQPGORIDLSNISBSVKNKERANDTAYGDTIPTYSFVRAHDSVQTV 573
Db 510 DNKLRLSLFLSLAKPLNQSGMNPJITNSL-VNRDDNAETAAPVSPSIFRAHDSVQDL 568
QY 574 IAKIVKEKIDTNSDGYTFTLDOLKDAFKIYNEDMAKVNKTYTHYINIPAYAILLSNMSV 633
Db 569 IADIITKAEINPNVVGYSFTMBEIKKAFEYLNKDLATEKKYTHYNTALSYALLLTKSSV 628
QY 634 PRVYVYDLYTDDGOYMAKSPYDAIATMLOGRIAYVSGGQSEEVHVKVNGNNOILSSVRY 693
Db 629 PRVYVYDLYTDDGOYMAKSPYDAIATMLOGRIAYVSGGQSEEVHVKVNGNNOILSSVRY 687
QY 694 GODLMSADDTQGTDLRSGLVTLVSNPNLNL-EGDSLTVNMGRANQARPLILGPK 752
Db 688 GKGALKATDT-GDRTTTSVAVIEGNNPSLRKASDRVVVNMGAHKNQAYRPLLLTTD 746
QY 753 DGVOYLYKDSD-TNIVKYTDANGNLTFTADDIKGYSTVDMSGYLAVVPVGA--DGQDV 809
Db 747 NGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKGYANPQVSGYLVGVVPVGAALIKMFAL 806
QY 810 RVAADTNQKADGKSLKTSNALSQVLYEGESNFQFANNADYTNKKTAENADFCKLGI 869
Db 807 RLARPHQQA---SVHQNAALDSRVMFEGSFQFATKKEBYTNVIAKNVDFKFAEWSV 863
QY 870 TSFEMAPQVVSATDGSFLDSIIQNGYAFSDRYDLAMSKNNKTKGSKDDLANALKALHANGI 929
Db 864 TDFENAPQVVSSTDGSFLDSVIQNGYAFTRDYLIGSKPNKYGTADDLVKALKALHSGKI 923
QY 930 QAIADWVPDQIYQVLPGEVVTAKRTNSYGNPFTDAYINNALLYATNTKSSGSIYQAOYGA 989
Db 924 KVMADWVPDQIYQVLPGEVVTAKRTNSYGNPFTDAYINNALLYATNTKSSGSIYQAOYGA 983
QY 990 ELDELKAKYPMFTVMNISTGKPIDPSTKIKOWEAKYFNGTNGVLGKAGYVLSDDATGY 1049
Db 984 FLUELQAKYPELFAKQKISTGVPMDFPSVKIKOMSAKYFNGTNGVLGKAGYVLSDDATGY 1043
QY 1050 FTVNENGD--FLPASFTGDQNAKTGFYDGTGMAYYSTSGNKAVNSFIYEGGHYIYFDK 1107
Db 1044 FNISDNKEINFLPKTL--NQDSQVGFSDYDGKGVYVYSTSGYQAKNTFISEGDKWYFDNN 1102
QY 1108 GHMVTGSYKAEDGNDYIFLPGIOMRDALYODAOQNSYVYGTGILYKGNWYPPFVDPNN 1167
Db 1103 GYMVTGA-QSINGVNYFLSNGQLRLDAILKNEDGTIAYYNGDGRYB-NGYQFQF--- 1156
QY 1168 ANKTVFRYEDANNVMAI GYRNMYGQTYFYDENGFOAKCOLLTDDKG-THYFEDENGAMAK 1226
Db 1157 --SGVYRHFN-NGEMSGLTVIDGQVQFDEMGGYQAKGFVTTADGKIRYFDKQSGNMVR 1213
QY 1227 NKXFV-NVGDDWYMDGNGNAVKGQPVVNNQILYFNPETGVQVKGQFITDAQGRTSYYDAN 1285
Db 1214 NREFIENEKGWLYLGEDGAATVGSQTINGQHLVFR-ANGVQVKGFEVTDHGRHSIYDGN 1272
QY 1286 SGALKSSGFFTPNGSDWYAEANGYVYKGFQVAENQDQWYVFDQTGQAKGAAKVQDGRD 1345
Db 1273 SGQDQIRNRF-----VRNAQGWQFVFD-NNGYAVTGTARTINGQL 1309
QY 1346 LYFENDPSGVQVKGDFATDESNGTTSFVHGDNQDKVGGFFPTGNNWYVADNNGNLVKGFQ 1405
Db 1310 LYFRA-NGVQVKGFEVTDYGRISYIDGNSGQOIRNFRVRNAQGWQFVFDNNGYAV---- 1364
QY 1406 EIDGKWHYFDEVTGQQAAGALVNGQQLYFVDSGTQVKGDFVTDGQGNSTSYDVNSGDK 1465
Db 1365 -----TGARTINGQHLYFRA-NGVQVKGFEVTDHGRHSIYDNGSGDQ 1406

QY 1466 KVGFFTTGNAMYADGQGNLAKGRKSIDNQDLYFDPATGKQVKGQLVSD-GRNYF 1524
 Db 1407 IRRFVRNAQGFYFNDNGYAVTGARTINGQLYF-RANGVQVKGFEVTDYGRISYD 1465

QY 1525 SGSGNMAK 1532
 Db 1466 ANSGRVR 1473

RESULT 13
 US-09-740-274-4
 ; Sequence 4, Application US/09740274
 ; Patent No. US20020031826A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Nichols, Scott E.
 ; TITLE OF INVENTION: Glucan-containing Compositions and Paper
 ; FILE REFERENCE: 0357CRD
 ; CURRENT APPLICATION NUMBER: US/09740, 274
 ; PRIOR FILING DATE: 2000-12-19
 ; PRIOR APPLICATION NUMBER: 09/210,361
 ; PRIOR FILING DATE: 1998-12-11
 ; PRIOR APPLICATION NUMBER: 09/007,999
 ; PRIOR FILING DATE: 1998-01-16
 ; PRIOR APPLICATION NUMBER: 08/478,704
 ; PRIOR FILING DATE: 1995-06-07
 ; PRIOR APPLICATION NUMBER: 09/009,620
 ; PRIOR FILING DATE: 1998-01-20
 ; PRIOR APPLICATION NUMBER: 08/485,243
 ; PRIOR FILING DATE: 1995-06-07
 ; PRIOR APPLICATION NUMBER: 09/008,172
 ; PRIOR FILING DATE: 1998-01-16
 ; PRIOR APPLICATION NUMBER: 08/482,711
 ; PRIOR FILING DATE: 1995-06-07
 ; NUMBER OF SEQ ID NOS: 6
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 4
 ; LENGTH: 1375
 ; TYPE: PRT
 ; ORGANISM: streptococcus mutans
 ; US-09-740-274-4

Query Match 38.4%; Score 3161.5; DB 3; Length 1375;
 Best Local Similarity 48.5%; Pred. No. 1.2e-158;
 Matches 662; Conservative 223; Mismatches 400; Indels 81; Gaps 31;

QY 1 MEKLLHYKLHKYKHWTTAVAS--IGLVSLVCA--GTVSAEDKVANDTTAQTAVGVDG 56
 Db 1 MEKKVRFKLKVKRWVTVSIAVAVTLTSLGSLVKADSTDDROQAVTESQASL----- 55

QY 57 QDOATTNDANTWT-TDITDADQANTNQDQAG-SDQSNQDQAKQDTANT-----DRNQ 108
 Db 56 ---VTTSEAKETLTAITDITSTATSATSQPTATVTDVNVSTNQSTNTTANTANFVVKPTTT 112

QY 109 ANSQTND-----NQATDQATS-----PATDGTSVQRDRANVAATAADQEG-----QTAP 153
 Db 113 SEQAKTDNSDKIITTSKAVNRLTATGKFPANNNTAHPKTVTDKIPIKPKIGLKQPS 172

QY 154 SEQEKBAAL-SLDNVKLDGKYVYVQADGSYKQNFATVNGQMLYFSDTSGALSTSTYS 212
 Db 173 LSQDDTAALGNKVRKNGKYYKKEDGTLOKYNALNTNGKTFPFD-ETGALSNNLTFS 231

QY 213 FSGQGTN--LVDDFFSHNKAYDSTAKSFELVNGYLTANSWYRPAGILRNGQTWEASNND 270
 Db 232 KKGNTNNDNTNSFAQYNQVYSTDVANFEVDHLYLTAESWYRPKYLLKDGKTWTQSTKD 291

QY 271 LRPVLMSWPKDQTVAYVYNNKYLASNETVETNETSQVDLNMKEAQSIQTKIEQKITS 330
 Db 292 FRPLMTWPDQETQRQYVYNNKQAQGIHQTYNT-ATSPLOLNLAQTIQTKIEKITA 350

QY 331 NSTQWLRTAMEAFVAAQPKWNSTEN-FNKGHLOGGALLYT-NSDLTPWANSYVLLNR 388
 Db 351 KNTNMLRQTIISAVKTSQAWNSDSEKPF--DHLQKGLLYSNNSKLTQANSYVLLNR 408

QY 389 TPTQDQGT--KYFTBEGGEGYFLLSNDVDNSNPVQVQAEQLNQLHLYLNMWDIVMGDKD 446
 Db 409 TPTNQTKDPRYATDRTTGGYFELLANDVDNSNPVQVQAEQLNQLHLYLNMWDIVMGDKD 468

QY 447 ANFDGVRVDAVDNADLLQVSNYFKNQYKVTDSSEANALAHISILSAWSLNDNQYNEDT 506
 Db 469 ANFDSIRVDAVDNVDADLLQIAGDYLKAAKGIHKNDKAAANDHLSILSAWSYNDPTYLHDD 528

QY 507 NGTALSIDNSSRLTSLAVLTQPGQRIDLSNLSESUNKERANDTAYGDTIPTYSFYRAH 566
 Db 529 GDNMINMDNRLRUSLLYSLAKPLNQSGMPLITNSL-VNRTDDNAETAAPVPSYFTRAH 587

QY 567 DSEVQTVIAKIVKEKIDTNSDGYFTFLDQDKDAFKIYNEDMAKVNKYTHYNTPAAYALL 626
 Db 588 DSEVQDLIRNIIRTEINPNVGVSYFTTEEIKKAFYINQDOLLATEKYKTHYNTALSALL 647

QY 627 LSNMESVPRVYGDLYTDGQYMAKSPYDAITATMLQRIAYVSGQSSEVHKVNNQ 686
 Db 648 LTNKSSVPRVYGDLYTDGQYMAKSPYDAITATMLQRIAYVSGQSSEVHKVNNQ 706

QY 687 ILSSVRYGQDLSADDTQGTDLRSRTSLVTLVSNLNDL--GGDSLTVNMGRAHANQAYR 745
 Db 707 IITSVRYGKALKATDT-GDRTTRTSGVAVIEGNNPSRLKASDRVVVNVNMGAAHKNQAYR 765

QY 746 PLILGTDGQVQSYLKDS--TNIIVKYTDANGNLTFTADDIKGYSTVDMSGYLAVWVPVGA 804
 Db 766 PLLLTDTNGIKAVHSDQEAAGLVRYTNDRGELIFTAADIKGYANPQVSGYLVWVPVGA 825

QY 805 DGQDVRVAADTNQKADGSKLTSALDSQVYIEGFSNFQDANNDADYTNKKIAENADFF 864
 Db 826 ADQDVRVAASTPSTDGKSVHQAALDSRVMPGFSNFQAFATKKEEYTVVIAKNVDF 885

QY 865 KKLGITSFEMAPQYVSATDGSFLDSIIQNGYAFSDRYDLAMSNNKYKSGKDDLANKAL 924
 Db 886 AEWGVTDFEMAPQYVSATDGSFLDSIIQNGYAFSDRYDLAMSNNKYKSGKDDLANKAL 945

QY 925 HANGIOAIADWVPDQIYQLPGEVVTAKRTNSYGNPTFDAYINNALLYATNTKSSGSDYQA 984
 Db 946 HSKGIKVMADWVPDQIYQLPGEVVTAKRTNSYGNPTFDAYINNALLYATNTKSSGSDYQA 1005

QY 985 QYGAFLDELKAKYPMFTVNMISTGKPIDPSTKIKOWEAKYPNGTNVLGKAGYVLSDD 1044
 Db 1006 KYGGAFLDELKAKYPMFTVNMISTGKPIDPSTKIKOWEAKYPNGTNVLGKAGYVLSDD 1065

QY 1045 ATGKYFTVNEGDFLPASFT----GDONAKTGPYDGTGMAYYSTSGNKAIVSFIYEGGH 1100
 Db 1066 ATNTYFSLVSDNTFLPKSLVNPNGHTSSSVTGLVPDQKGYVYVYSTSGNKAIVSFIYEGGH 1125

QY 1101 YYYFDKDHMTGSKYKAEEDNDYFPLPNGIQMRDAIYQDAQNSYYVYGRGILYK----- 1155
 Db 1126 WYFDFNNGYVMTGA-QSINGANYFSLNGIQLBNALYDNGNKVLSYDNGDRYENGYYL 1184

QY 1156 -GDNWYFVDPNNANKTVRYFDANNVMAIGYRNMYGQTYFYFDENGFOAKGOLLTDKG- 1213
 Db 1185 FGQW-----RYFQ-NGIMAVGLTRVHGAQYVFDASGFOAKGQFITADGK 1229

QY 1214 THYFEDNCGAMAKNFV-NVGDDWYMDGNGNAVKQYVNNQILYFNBPETGVQVKGQFI 1272
 Db 1230 LRYFDRDSNQISNRPVRNKGWFLFDHNGVAVTGTFTNGQRLYFKS-NGVQAKGEFI 1288

QY 1273 TDAQRTSYDANSALKSSGFTTPNGSDWY-YAENGYYVYKPKQV 1317
 Db 1289 RDANGVLYRYYDPSNGNEVRNFRVRNKGWFLFDHNGIATGARVV 1334

RESULT 14
 US-10-383-930-35
 ; Sequence 35, Application US/10383930
 ; Publication No. US20040127400A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Smith, Daniel J
 ; APPLICANT: Taubman, Martin A

; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 35
; LENGTH: 1375
; TYPE: PR1
; ORGANISM: Streptococcus mutans
; US-10-383-930-35

Query Match 38.4%; Score 3161.5; DB 4; Length 1375;
Best Local Similarity 48.5%; Pred. No. 1.2e-158;
Matches 662; Conservative 223; Mismatches 400; Indels 81; Gaps 31;

QY 1 MEKKLHYKLHKVKKHWVTIAVAS--IGLSVLGA--GTVSAEDKVANDTTAQATVGDVGTG 56
DB 1 MEKKVRFKLKVKVKKVWTVSIASAVTLTSLGSLVKADSTDDRQOAVTESQASL----- 55
QY 57 QDOATTNDANTWT-TDTDADQSANTNQOAG--SDQSNQDQAKQDNTANT-----DRNQ 108
DB 56 ---VTTSEAAKETLTATDTSTATSQPTATVTDNVSTTNQSTNTTANTANFVVKPTTT 112
QY 109 ADNSQTDN-----NOATDOATS-----PATDGTSGVQRDDAANVATAADQEG---QTAP 153
DB 113 SEQAKTDNSDKIITTSKANVRLTATGKFPANNNTAHPKVTVDKIIVPKIKGLKQPS 172
QY 154 SEQEKAAL-SLDNVKILDGKYYVQADGSKNFATVNGQMLYFSDTSGALSSSTYS 212
DB 173 LSQDDIAALGNVKNIRKNGKYYKYKEDGTQKNYALNNGKTFPFDD-ETGALSNNTPS 231
QY 213 FSGQITN--LVDDFSHNKAYDSTAKSPFLVNGYLTANSWYRPAGILRNGQITWEASND 270
DB 232 KKGNIITNNNTNSFAQYNQVYSTDVANFEHVDHYLTAESWYRPKYILKDGKWTQSTEKD 291
QY 271 LREVLMSWPKDQTVAYVYNNMKYLSANETVNETSVDNLNKEAQSIOTKIEOKITSD 330
DB 292 PRELLMTWFPDQETQRYVYNNMAQLGIHQTYNT-ATSPLOLNLAQOTQTKIEEKITAE 350
QY 331 NSTQWLRTAMEAFVAAQPKWNMSTEN-FNKGDLQGGALLYT-NSDLTPMANSDYRLNLR 388
DB 351 KNTNWLRLQITSAFVKTSQAWNSDEKPFDD--DHLQKGLLYSNNSKLTQANSNYRILNR 408
QY 389 TPTQOQDGTK--KYFTEGGEGYEFLISNDVNSNPVVAEQLNLHLYLNNWGDIVMGDKD 446
DB 409 TPTNQTGKDPRTYDRTTIGGYEFLLANDVDNSNPVVAEQLNLHLYLNNWGDIVMGDKD 468
QY 447 ANFDGVRVDVAVNVDNADLLQVYNSYKDNVKTVDSEANALAHISILEAWSLNDQVNEPT 506
DB 469 ANFDSIRVADVNDVDADLLQIADGYLKAAGIHKNDKAANDHLISLEAWSYNDTPYLHDD 528
QY 507 NGTALSIDNSSRLTSLAVLTQPGQRIDLSNLISSEVKNKERANDTAYGDTIPTYSFVRAH 566
DB 529 GDMNINNDNRLLSLLYSLAKPLNQRSGMNPFLITNSL-VNRTDDNAETAAPVSYSFIRAH 587
QY 567 DSEVQTVIAKIVKEKIDTNSDGYTFTLDQLKDAFKLYNEDMAKVNKTYTHYNIPAAVALL 626
DB 588 DSEVQDLIRNIIRTEINPNVVGYSPTTEEIKKAFEYNKDLLATEKKYKTHYNITALSAYLL 647
QY 627 LSNMESVPRVYVYGDLYTDDQYMAKSPYVDAITMLQGRIAVYSGQSEVHKVNGNQ 686
DB 648 LTNKSSVPRVYVYGDVMTDDQYMAKHTINYEALETLLKARIKTVSGQAMRNOQV-GNSE 706
QY 687 ILSSVRYGDLMSADDTQGTDLRSRTSLGLVTLVNSDNLNLDI--GGDSLTVMNGRAHQAQYR 745
DB 707 IITSVRYGKALKATDT-GDRTTTSVGVAVIEGNPNPSRLKASDRVVVNNMGAHKKQAQYR 765
QY 746 PLILGTGDKGVQSLKQSD--TNIVKYTDANGNLTFDADDIKGYSTVDSMGVLAWVVPVGA 804

DB 766 PLLITTDNGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKGYANPQVSGYLVWVPVGA 825
QY 805 DGQDVRVAADTNQKADGSKLTKTSAALDSQVIYEGSFNFQDFAANNADATYNTKKIAENADFF 864
DB 826 ADQDVRVAASTAPSTDGKSVHQNAALDSRVMEPEGSFNOAFATKKEBYTNVVIKAVDKF 885
QY 865 KKLGLITSFEMAPQYYSATDGSFLDSIQNGYAFSDRYDLAMSKNNKYGSKDDLAKALKAL 924
DB 886 AEWGVTDFEMAPQYYSSTDGSFLDSVIQNGYAFDTRDYLGISPKNKYGTADDLVLKALKAL 945
QY 925 HANGIOALADWVPDQIYOLPGEEVVTAKRTNSYGNPTFDAYINNALLYATNTKSSGSDYO 984
DB 946 HSKGILKWADWVPDQMYALPEKEVVTATRVKGYTFVAGSQIKNTLYVVDGSSGKDDQA 1005
QY 985 QYGGAFLELDELKAKYDPMFTVMNMISTGKPIDPSTKIKQWEAKYFNGTNVLGKAGYVLSDD 1044
DB 1006 KYGGAFLBELQAKYPELPARKQISTGVPMDPSPVKIKQWSAKYFNGTNILGRGAGYVLKQD 1065
QY 1045 ATGKYFTVNEGDPLPASFT----GDQNAKTGFYDGTGCTMAYYSTSGNKAVNSFIYEGGH 1100
DB 1066 ATNTYFSLVSDNTFLPKSLVNPNNHGTSSSVTGLVFDGKGYVYVYSTSGNQAKNAFTSLGN 1125
QY 1101 YVYFDKDGHWMTGSKAEDGNDYFPLNGIOMRDAIYQDAQNSVYVYGRGTGLYK----- 1155
DB 1126 WYTFDNGWVMTGA-QSINGANYFSLNGIQIRNALIYDNGNKVLSYNGNDRRYENGYYL 1184
QY 1156 -GDNWYFPVDPNNANKTVRYFDANNVMAIGYRNMYGQTYFYFDENGFOAKGQLLTDKG- 1213
DB 1185 FGOQM-----RYFO-NGIMAVGLTRVHGAVQYFDFASGFOAKGQFITTADGK 1229
QY 1214 THYFEDNCGAMAKNFV-NVGDDWYMDGNGNAVKQYVNNQILLYFNSETGVQVKGQFI 1272
DB 1230 LRYFDRDSNQISNRFVRNSKGWELFDHNGVAVTGTVTFNQORLYFKP-NGVQAKGEFI 1288
QY 1273 TDAQRTSYVDANSALKSSGFFTPNGSDWY-VAENGYYVYKGFQV 1317
DB 1289 DRANGYLRYDPSNGEVRNFRVRNSKGWELFDHNGIATVGARVV 1334

RESULT 15
US-10-797-821-35
; Sequence 35, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 35
; LENGTH: 1375
; TYPE: PR1
; ORGANISM: Streptococcus mutans
; US-10-797-821-35

Query Match 38.4%; Score 3161.5; DB 5; Length 1375;
Best Local Similarity 48.5%; Pred. No. 1.2e-158;

THIS PAGE BLANK (USPTO)

GenCore version S.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 20:42:47 ; Search time 8.84317 Seconds
(without alignments)
2306.008 Million cell updates/sec

Title: US-10-797-821-38

Perfect score: 8237

Sequence: 1 MEKKLHYKLHKVKKHWTTA.....FVRIGQWYFGNDGAATNL 1554

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 97014 seqs, 1312538 residues

Total number of hits satisfying chosen parameters: 97014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA_New.*

- 1: /cgn2_6/prodata1/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/prodata1/pubpaa/US06_NEW_PUB.pep.*
- 3: /cgn2_6/prodata1/pubpaa/US07_NEW_PUB.pep.*
- 4: /cgn2_6/prodata1/pubpaa/PCT_NEW_PUB.pep.*
- 5: /cgn2_6/prodata1/pubpaa/US05_NEW_PUB.pep.*
- 6: /cgn2_6/prodata1/pubpaa/US10_NEW_PUB.pep.*
- 7: /cgn2_6/prodata1/pubpaa/US11_NEW_PUB.pep.*
- 8: /cgn2_6/prodata1/pubpaa/US60_NEW_PUB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	545.5	6.6	2710	US-11-051-453-41	Sequence 41, Appl
2	443.5	5.4	2367	US-11-051-453-42	Sequence 42, Appl
3	363	4.4	1992	US-11-013-759-3	Sequence 3, Appli
4	363	4.4	1992	US-11-013-759-13	Sequence 13, Appl
5	363	4.4	2047	US-11-013-759-4	Sequence 4, Appli
6	363	4.4	2047	US-11-013-759-9	Sequence 7, Appli
7	341.5	4.1	2053	US-11-013-759-9	Sequence 9, Appli
8	306.5	3.7	2314	US-11-013-759-11	Sequence 11, Appl
9	302.5	3.7	396	US-11-022-562-228	Sequence 228, App
10	290	3.5	726	US-11-052-554A-217	Sequence 217, App
11	286.5	3.5	1647	US-11-052-554A-260	Sequence 260, App
12	276.5	3.4	2399	US-11-052-554A-92	Sequence 92, Appl
13	267	3.2	338	US-11-052-554A-228	Sequence 228, App
14	267	3.2	1155	US-10-793-626-1780	Sequence 1780, Ap
15	265	3.2	1767	US-11-052-554A-372	Sequence 372, App
16	263	3.2	3194	US-11-052-554A-90	Sequence 90, Appl
17	261.5	3.2	744	US-10-873-528-184	Sequence 184, App
18	258	3.1	2902	US-11-052-554A-91	Sequence 91, Appl
19	256.5	3.1	677	US-10-873-528-155	Sequence 155, App
20	256	3.1	619	US-11-052-554A-229	Sequence 229, App
21	256	3.1	701	US-11-052-554A-231	Sequence 231, App
22	255	3.1	658	US-10-873-528-17	Sequence 17, Appl
23	251	3.0	5291	US-11-052-554A-281	Sequence 281, App
24	249	3.0	1588	US-11-052-554A-280	Sequence 280, App
25	246.5	3.0	690	US-11-052-554A-232	Sequence 232, App

ALIGNMENTS

RESULT 1

US-11-051-453-41
; Sequence 41, Application US/11051453
; Publication No. US20050287150A1
; GENERAL INFORMATION:
; APPLICANT: AMBROSINO, DONNA
; APPLICANT: BABCOCK, GREGORY J.
; APPLICANT: BROERING, THERESA
; APPLICANT: GRAZIANO, ROBERT
; APPLICANT: HERNANDEZ, HECTOR JAVIER
; APPLICANT: LOWY, ISRAEL
; APPLICANT: MANDELL, ROBERT
; APPLICANT: MOLRINE, DEBORAH
; APPLICANT: THOMAS, JR., WILLIAM D.
; APPLICANT: ZHANG, HUI-FEN
; TITLE OF INVENTION: ANTIBODIES AGAINST CLOSTRIDIUM DIFFICILE TOXINS AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: MJ1-001
; CURRENT APPLICATION NUMBER: US/11/051,453
; CURRENT FILING DATE: 2005-02-04
; PRIOR APPLICATION NUMBER: 60/542,357
; PRIOR FILING DATE: 2004-02-06
; PRIOR APPLICATION NUMBER: 60/613,854
; PRIOR FILING DATE: 2004-09-28
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 41
; LENGTH: 2710
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-11-051-453-41

Query Match 6.6%; Score 545.5; DB 7; Length 2710;
Best Local Similarity 20.1%; Pred. No. 1.2e-20;
Matches 380; Conservative 226; Mismatches 624; Indels 661; Gaps 91;

QY 35 VSAEDKAVNDTTAQATVGTGQDQATTNDANTTTDTTADASANTNODQAG-----S 88
Db 913 VETKEIFSKYSEHITKEITKNSIIT-DVGNLLDNLQLDHTSOVTLNAAFFQSLI 971
QY 89 DQSNQD-----QAQDQDANT--DRNQADNSQTDNNQAQDQATSPATDGT 131
Db 972 DYSSNKQVLDLSTSVKVLQAFLSTGLTIYDLSQLVNLSNAVNDINVLPTTEGI 1031
QY 132 SVQRR--DAANVATA----AQEGQTAPSEQKSAALSNDVKL-IDGKYVYVQADGSYK 184
Db 1032 PIVSTILDGINLGAIAKELLDEHPLKKLEAKVGLVAINMNSLSAATVASTIVGIGAEV 1091

Sequence 141, App
Sequence 2, Appli
Sequence 1, Appli
Sequence 190, App
Sequence 368, App
Sequence 2964, Ap
Sequence 1, Appli
Sequence 212, App
Sequence 188, App
Sequence 185, App
Sequence 282, App
Sequence 13, Appl
Sequence 45, Appl
Sequence 127, App
Sequence 283, App
Sequence 191, App
Sequence 74, Appl
Sequence 216, App

26 245.5 3.0 1290 6 US-10-485-517-141
27 241.5 2.9 485 6 US-10-630-203-2
28 241.5 2.9 485 7 US-11-103-037-1
29 241.5 2.9 485 7 US-11-195-538-1
30 240 2.9 332 6 US-10-873-528-190
31 235 2.9 1571 7 US-11-052-554A-2
32 228.5 2.8 2551 7 US-11-052-554A-368
33 227.5 2.8 5024 6 US-10-793-626-2964
34 226.5 2.7 1268 7 US-11-052-554A-1
35 224.5 2.7 1448 6 US-10-485-517-212
36 223 2.7 340 6 US-10-873-528-188
37 220 2.7 693 6 US-10-873-528-185
38 219.5 2.7 1345 7 US-11-052-554A-282
39 216 2.6 1327 7 US-11-052-554A-13
40 216 2.6 1889 7 US-11-102-476-46
41 212 2.6 1263 6 US-10-485-517-127
42 210 2.5 1461 7 US-11-052-554A-283
43 209 2.5 627 6 US-10-873-528-191
44 208 2.5 1484 6 US-10-517-939-74
45 206.5 2.5 567 6 US-10-485-517-216

Qy	185	KNFALTVNGQMLY	PDSTGAL	STSTVSF	QCGTTLNV	DDFSSHNKAYD	STAKSFELVN	242		
		: :	: :	: :	: :	: :	: :			
Db	1092	TIFLLPIAG	-----	ISAGIPSLVNN	ELIHDKA	-----	TSVYNY	1125		
		: :	: :	: :	: :	: :	: :			
Qy	243	GYLTSANWYRPAGI	LRNGQTWEAS	NENDLR	PVLMWMPDKD	TOQAVVYMNKYL	-----	296		
		: :	: :	: :	: :	: :	: :			
Db	1126	FNLHSEKKYGP	-----	LKTEDD	--KILV--	PIDDLVISBIDFNNNSIKL	GTC	1169		
		: :	: :	: :	: :	: :	: :			
Qy	297	-----	1-SANETEVTN	-----	ETSQVDLNKEAOS	IOTKIE	324			
		: :	: :	: :	: :	: :	: :			
Db	1170	NILAMEGSGH	TVTGNIDHPFSSPS	ISSHIPSL	SIYSAIGIETENL	DFSK	-----	KIM 1222		
		: :	: :	: :	: :	: :	: :			
Qy	325	QKITSDNSTOWL	RAMEAFVAAQPKNM	STENFNKG	DHLOGGALLYTNS	DLTP	-----	377		
		: :	: :	: :	: :	: :	: :			
Db	1223	MLPNAPS	RVFWWET	-----	CAVPGLR	-SLEN	-----	DGTRLLSDIRLYPGKFYWR 1268		
		: :	: :	: :	: :	: :	: :			
Qy	378	WANSDYRLN	RPTQODGTYKY	FTEGEGGYE	FLLSNDVDNS	---	PVVOAQL	NOLHY 433		
		: :	: :	: :	: :	: :	: :			
Db	1269	YAFPDYAIT	TLKPVEYDNIK	-----	IKLKDTRNFIMPT	ITNBEIRNKLSY	1315			
		: :	: :	: :	: :	: :	: :			
Qy	434	LMNWGDIW	MGDKDANFGVR	VDADVNDV	NADLLQVYS	-----	NYFKONYKVTDSEANALAH 488			
		: :	: :	: :	: :	: :	: :			
Db	1316	-----	SPDG	-----	AGTYISLLSSYPISTIN	LSKD	-----	1342		
		: :	: :	: :	: :	: :	: :			
Qy	489	ISILEAWSLN	-DNOYNEDTNG	TALSDNS	SRLTS	LAVLTKOPQOR	IDLNSL	ISESVNKR 547		
		: :	: :	: :	: :	: :	: :			
Db	1343	---DLWIF	INIDNEVRE	-----	ISIENGT	-IKKGKLIKDVLSKID	-----	INK- 1380		
		: :	: :	: :	: :	: :	: :			
Qy	548	ANDTAYGDT	PTYSPVRAH	SEVOTVI	AKIVKEKIDTNS	DGYFTLDOL	KDAPKI	YNEDM 607		
		: :	: :	: :	: :	: :	: :			
Db	1381	NKLIIGN	QNOTDFS	-----	GDID	-NKDRIYFLT	CELDDKISL	---I 1416		
		: :	: :	: :	: :	: :	: :			
Qy	608	AKVNKTYTHYNI	PAAYALLSN	MESVPRVY	YGDLYTDG	GVMAKKS	PPYDAI	ATMLQRI 667		
		: :	: :	: :	: :	: :	: :			
Db	1417	IEIML	-----	VAKSYLLLS	-----	-----	GDK	1434		
		: :	: :	: :	: :	: :	: :			
Qy	668	AYVSGOGSE	EVHKVNGN	NNQILSS	VRYGQDLMS	ADDTQGTDL	SRTSGL	VLTVSNDPNLDLG 727		
		: :	: :	: :	: :	: :	: :			
Db	1435	NYLISNL	SNTIEKIN	-----	-----	TLGLDSKNI	AYNTVDES	N	1472	
		: :	: :	: :	: :	: :	: :			
Qy	728	GDSLTVM	GRAHANQVR	PLILGT	KOGVQSVLK	SDTNI	VKYTDAN	---GNLFT	ADDDIK 784	
		: :	: :	: :	: :	: :	: :			
Db	1473	AISRTS	QKSIIH	-----	YKKDSKN	ILEFYNDST	LEFNSK	DIFAE	DIN 1514	
		: :	: :	: :	: :	: :	: :			
Qy	785	GYSTVDMS	GYLAVWVP	CAKGQD	QVRVAADT	NQADGKSL	KTSAALDS	---QVI	YEG--- 838	
		: :	: :	: :	: :	: :	: :			
Db	1515	VFMKDD	QINTITG	-----	KYVDNN	-----	TKSDIFS	LSLVSKNQK	VNGLYLN 1558	
		: :	: :	: :	: :	: :	: :			
Qy	839	---FSNFQD	FANNADATY	NKK	-----	IAENADFF	PKLGIT	SFE	MAPQY---VSAT	DGSF 886
		: :	: :	: :	: :	: :	: :			
Db	1559	ESVYSSYL	DFVKNSDGH	HTNSFN	WNFLDNI	SPWKLFG	ENIN	FVIDKY	FTLVGK	TNLGY 1618
		: :	: :	: :	: :	: :	: :			
Qy	887	LDSIIQNG	YAFSDRY	DLAMSK	WNKYGSK	DDLANAK	HALHANG	IOAIAD	---WVP	QIYQLP 944
		: :	: :	: :	: :	: :	: :			
Db	1619	VEFICDN	---NKNID	IYFG	EWKTS	SSKSTI	-----	FSNGRNV	VVEPI	YNPD-----T 1663
		: :	: :	: :	: :	: :	: :			
Qy	945	GEVVTA	---KRTNS	YGNPT	DAYINN	ALVATNT	KSS	-----	GSDYQAO	YGGAFLDELKAK 997
		: :	: :	: :	: :	: :	: :			
Db	1664	GEDIST	SLDFS	YEPLYG	---IDRY	INKVL	APDLYT	SLININT	NY---YSNEY	

; ORGANISM: Clostridium difficile.
US-11-051-453-42

Query Match 5.4%; Score 443.5; DB 7; Length 2367;
Best Local Similarity 18.3%; Pred. No. 2e-15;
Matches 360; Conservative 263; Mismatches 604; Indels 739; Gaps 87;

QY 140 NVATAADQEGQTAPSEQKSAALSLDNVLIIDGKYYVQADGSKYKKNFAITVNGQMLYFD 199
D 140 NVATAADQEGQTAPSEQKSAALSLDNVLIIDGKYYVQADGSKYKKNFAITVNGQMLYFD 199
D 508 NISQATEQKNSLWQFEERAKIQFE-----EYKKN-----YFE 541
QY 200 SDTGALSSSTYGFSGTNN-----LVDDPSSHKNKAYDSTAKSELVNGY----- 244
D 200 SDTGALSSSTYGFSGTNN-----LVDDPSSHKNKAYDSTAKSELVNGY----- 244
D 542 ---GALGEDNLQFSQNTVTDKEYLLEKISSSTKSS-----GGYVHYIVQLQGD 589
QY 245 ---LTANSWYRPAGLRNGQTWEASNENDLRPLVMSWNPDKDTQVAYVYNNKYL 296
D 245 ---LTANSWYRPAGLRNGQTWEASNENDLRPLVMSWNPDKDTQVAYVYNNKYL 296
D 590 ISYEACNLFAKNPYDSILFORN-----IEDSEVAY-----YY 622
QY 297 SANETEVTNETSQVDLNKEAQSIQTKIEQKIT-----SDNSTQWLRT 338
D 297 SANETEVTNETSQVDLNKEAQSIQTKIEQKIT-----SDNSTQWLRT 338
D 623 NPTDSEI---QBIDKYRIPDRISDRPKIKLTFIGHGKAEFNTDIPAGLDVDSLSSEIET 678
QY 339 AME-AFVAAPK-----WNSTENFNKGDHLOGGALLYTN---SDLTPWANSYRLL 386
D 339 AME-AFVAAPK-----WNSTENFNKGDHLOGGALLYTN---SDLTPWANSYRLL 386
D 679 AIGLAKEDISPKEIEINLLGCNMFYSVNVETYPGKLLRVKDKVSELMPMSQDSIIV 738
QY 387 NRPTPTO---QDCTKKYFTGGG-----GGEFLLSDNDVNSNPVVOAEOQLNQ 430
D 387 NRPTPTO---QDCTKKYFTGGG-----GGEFLLSDNDVNSNPVVOAEOQLNQ 430
D 739 SANQYEVRISEGRRELLDHSGEWINKESIIKDISEKYSISFNPKEN-KIIVKSKNLP 797
QY 431 LHYLM-----NWGDIVMGDK-----DANFDGVRVDAVDNVDNADLLQVYS 469
D 431 LHYLM-----NWGDIVMGDK-----DANFDGVRVDAVDNVDNADLLQVYS 469
D 798 LSTLLQEIRNNSSDIELEKVMACEINVISNIETQVVEERIEEAKSLTSDSI----- 853
QY 470 NTFKONYKVTDSANALAHISLEAWSLNDQVNETDNGTALSIDNSSRLT-----SLAVLT 526
D 470 NTFKONYKVTDSANALAHISLEAWSLNDQVNETDNGTALSIDNSSRLT-----SLAVLT 526
D 854 NYIKNEFKLIESEALCDL-----KQNELEDHSFISFEDISETDEGFSIRFIN 903
QY 527 KOPGORI-----DLSNLISESVNKERAN--DTAYGDTIPTYSFVRANDSEVQTVI 574
D 527 KOPGORI-----DLSNLISESVNKERAN--DTAYGDTIPTYSFVRANDSEVQTVI 574
D 904 KETGESIFVETEKTIIFSEYANHITEISIKIGIFDVNGKVKVKNLDTTH--EYNTLN 961
QY 575 AKI-----VKEKIDTNSDGYFTLQLOKDAFIYN----- 604
D 575 AKI-----VKEKIDTNSDGYFTLQLOKDAFIYN----- 604
D 962 AAFPIQSLIEYSSKESLSNLSVAMKVQVYAQLFSTGLNTITDAKVVSLVSTALDETID 1021
QY 605 -----EDMAKVNTYTHYINIPAAAYALLSNMES----- 632
D 605 -----EDMAKVNTYTHYINIPAAAYALLSNMES----- 632
D 1022 LLPTLSEGLPIIATIIDGVSGLGAAIKELSETSDPLLQRIEAKIGIMAVNLTTATTAIIT 1081
QY 633 -----VPRVYGDLYTDDGQYMAKSPY-----DAIATM 662
D 633 -----VPRVYGDLYTDDGQYMAKSPY-----DAIATM 662
D 1082 SLSGIASGSIILLVPLAGISAGIPLSVNVELVRD--KATKVVDYFKHVSIVETEGVFTL 1139
QY 663 LQCRITAYSGGQSE--VHKVNGNNOIL-----SSVR 692
D 663 LQCRITAYSGGQSE--VHKVNGNNOIL-----SSVR 692
D 1140 LDDKVMW---QODDLVISEIDPNNSNIVLGKEIWEHMEGGSGHTVTDIDHFFSAPSIT 1195
QY 693 YGQDLMSADD---TQGTDLRSRTSLGLVTLVSNPDN-----LDLGGDSLTYN 734
D 693 YGQDLMSADD---TQGTDLRSRTSLGLVTLVSNPDN-----LDLGGDSLTYN 734
D 1196 YREPHLSIYDVLEVQKEELDLSKDLMLV-PNAPNRVFAWETGTWGLRSLENDGTCLLDR 1254
QY 735 MGRAHANOAYRPLILGTGQVQSYLKD--SDTNVYKVTDANGNL-----TFTADDIK--G 785
D 735 MGRAHANOAYRPLILGTGQVQSYLKD--SDTNVYKVTDANGNL-----TFTADDIK--G 785
D 1255 IRDNEGEFERYWFAPFIADALITTLKPRYEDTNIRINLSNTRSFIVPIITTEYIREKLS 1314
QY 786 YSTVDMSGYLA-----VWVPVGAKDGQDVRVAADTNOKADGSKLS 827
D 786 YSTVDMSGYLA-----VWVPVGAKDGQDVRVAADTNOKADGSKLS 827
D 1315 YFYSGGGTYALPLSYNMGINIELESVDWIIDVNVVDVTIESDKIKGD----- 1367
QY 828 AALDSQVITYEGFSNFQDPANNDADYTNKKIAENADPFKKLGITSEFMAPQYVSATDGSFL 887
D 828 AALDSQVITYEGFSNFQDPANNDADYTNKKIAENADPFKKLGITSEFMAPQYVSATDGSFL 887

RESULT 3

US-11-013-759-3

; Sequence 3, Application US/11013759

; Publication NO. US20050249747A1

; GENERAL INFORMATION:

; APPLICANT: Loosmore, Sheena M.

Db 1368 -----LIEGILSTLSIEENKIILNSHEINFSGEVNSGFGVSLTF-----SILEG-I 1413
QY 888 DSIIONGYAFSDRYDLAMSKNNKYKSGKDDLANAKALKAHANGIOAIADM-----VP 937
D 888 DSIIONGYAFSDRYDLAMSKNNKYKSGKDDLANAKALKAHANGIOAIADM-----VP 937
D 1414 NALIE-----VDL-LSKSYKLLISGEL--KILMLNSNHIQOKDIYIGFNSBELQKNIP 1462
QY 938 DQYQYLPGEVWVTAKTNSYGNPTFPDAYINNALYATNTKSSGSDYQAOYAGAFDELKAK 997
D 938 DQYQYLPGEVWVTAKTNSYGNPTFPDAYINNALYATNTKSSGSDYQAOYAGAFDELKAK 997
D 1463 YSPVDSGEKE-----NGFING-----STKE-----GLFVSEL--- 1489
QY 998 YPDMFTVNMI--STGKPIDP--STKIKQWEAKYFNGTNVLKAGAGYVLSDD----- 1044
D 998 YPDMFTVNMI--STGKPIDP--STKIKQWEAKYFNGTNVLKAGAGYVLSDD----- 1044
D 1490 -PDVLIISKYMDSDSKSPFSGYYSNNLKDVKITKDNVNL---TGYLKKDDIKISLSLTL 1545
QY 1045 ---ATCKYFTV--NENG-----DPLPASFTCDQNAKT-----G 1072
D 1045 ---ATCKYFTV--NENG-----DPLPASFTCDQNAKT-----G 1072
D 1546 QDEKTIKLSVHLDSEGVAEILKFMKRGKSTNTSDSL-MSFLESMAIKSIFVNFPLQSNIK 1604
QY 1073 FYYDGTGMAYYTSNGKAVNSFIY---EGGH-----YYFYDKDGHMVTG 1113
D 1073 FYYDGTGMAYYTSNGKAVNSFIY---EGGH-----YYFYDKDGHMVTG 1113
D 1605 FILD---ANFIISGTSIGQPEFICDENNNIQPYFIKFWLTETNTLYVGNRQNMIVEP 1660
QY 1114 SYKAEDCND-----YYP---LPGIOMRDALITQDAQNSYYYGRTGILYKGDN 1158
D 1114 SYKAEDCND-----YYP---LPGIOMRDALITQDAQNSYYYGRTGILYKGDN 1158
D 1661 NYDLDDSGDISSTVINFSQKLYLGIDSCVKNKVISPNIVTDEINIT-----PVYETNN 1713
QY 1159 WYP-----FV 1163
D 1159 WYP-----FV 1163
D 1714 TPEVIVLDANYINEKINVININDLSIRYVWSNDGNDPILMSTSEENKVSQVKIRFVNVFK 1773
QY 1164 DPNNANKTVFRYFDANNVMA-----IGYR-----NMVQOTYYFDENGFOA 1203
D 1164 DPNNANKTVFRYFDANNVMA-----IGYR-----NMVQOTYYFDENGFOA 1203
D 1774 DXTLANKLSFNSDKQDVPVSEIILSFTPSYVEDGLIYDGLVLSLYNEKFIYNNFGMV 1833
QY 1204 KQOLLTDDGTHYFDEEDGAMAKNFVNVGDDWYMDG-NGNAV-KQYVNNQILYFNP 1261
D 1204 KQOLLTDDGTHYFDEEDGAMAKNFVNVGDDWYMDG-NGNAV-KQYVNNQILYFNP 1261
D 1834 SGLIYIND-SLYYFKPPVNNLITG-FVTVGDDKYYFNPINGGAASIGETIIDDKNYYFN- 1890
QY 1262 ETGVQVKGFITD-----AQGRTSYDANSALKSSGFTPNGSDWYAEVGVYKGF 1315
D 1262 ETGVQVKGFITD-----AQGRTSYDANSALKSSGFTPNGSDWYAEVGVYKGF 1315
D 1891 QSGVLQTVFSTEDGKYPAPANTLDENLEGEAIDFTGKLIIDENIYYFEDN---YRGAV 1947
QY 1316 QVAENODQWYFDTQTTQKQAKGAAKVDGRDLVFNPDGSGVQVKGDFAT-----DSG-- 1366
D 1316 QVAENODQWYFDTQTTQKQAKGAAKVDGRDLVFNPDGSGVQVKGDFAT-----DSG-- 1366
D 1948 ENKELDGEHMFSPETGKAFKGLNQIGDDKYFNSD-GVMQGFVINDKNKHVFDGSGVM 2006
QY 1367 -----NTSYHGDGDKVVGPPFTG-----NNA 1390
D 1367 -----NTSYHGDGDKVVGPPFTG-----NNA 1390
D 2007 KVGYTEIDGKHFPYFAENGEMQIGVFNTEDGKIFYAHNNEDLGNEEGEEISYSIGILNPNK 2066
QY 1391 WYVADNNGNLVKGFOEI-DGKWHFDEVTGQQAAGALVNGQOLYFD----- 1436
D 1391 WYVADNNGNLVKGFOEI-DGKWHFDEVTGQQAAGALVNGQOLYFD----- 1436
D 2067 IYFDDSFYAVVWGWKDLBGSKYFDEDTAEAYIGLSLINDGQYFNDGIMQVGFVTIN 2126
QY 1437 -----VDSGIQVKGD--FVTDGOG--NTSYDVSNSGDK----- 1465
D 1437 -----VDSGIQVKGD--FVTDGOG--NTSYDVSNSGDK----- 1465
D 2127 DKVYFVDSGIIESGVQNDNNFYDDNGIYQIGVDFDSDGKYKYPAPANTVNDNIYGOA 2186
QY 1466 -KVGFFFTGDNAWYADQGNLAKGRKSIDNOD--LYFDPATGKOVKG----- 1511
D 1466 -KVGFFFTGDNAWYADQGNLAKGRKSIDNOD--LYFDPATGKOVKG----- 1511
D 2187 VEYSGLVRVGEDVYFGETYTTIETGWIYDMENESDKYFVFPETKACKGINLIDDIKYFV 2246
QY 1512 -----QLVSDIGRNNYFDSGSGNMMAKRVFVRIGDQWYIFGNDG 1549
D 1512 -----QLVSDIGRNNYFDSGSGNMMAKRVFVRIGDQWYIFGNDG 1549
D 2247 DEKIMRTGLISFENNNYFNE-NGEI-QFGYINIEDKMFYFGE 2290

RESULT 4

US-11-013-759-13
; Sequence 13, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loomis, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN OF MORAXELLA
; FILE REFERENCE: 1038-921MIS:jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patentin ver. 2.1
; SEQ ID NO 13
; LENGTH: 1992
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-13

Query Match 4.4%; Score 363; DB 7; Length 1992;
Best Local Similarity 20.6%; Pred. No. 2.5e-11;
Matches 377; Conservative 196; Mismatches 682; Indels 578; Gaps 86;

QY 18 TIAVASIGLVSLVAGTVAEDKANDTTAAQTAVGVDTGQDQATTNDANTN-----TTDD 73
DB TIAVASIGLVSLVAGTVAEDKANDTTAAQTAVGVDTGQDQATTNDANTN-----TTDD 73
QY 196 TIAIGSDATSSSIAIGALG-----GTRAQLOQSIALGQSVVTSQDNNRRPAYTPNTQ 249
DB TIAIGSDATSSSIAIGALG-----GTRAQLOQSIALGQSVVTSQDNNRRPAYTPNTQ 249
QY 74 TADQSAN-TNODQAGSDQSNODQADQDANTDRNQADNSQTDNNQATQATSPATDGT 132
DB TADQSAN-TNODQAGSDQSNODQADQDANTDRNQADNSQTDNNQATQATSPATDGT 132
QY 250 ALDPKQATNTKAG-PLSIGNSIKRKIN-----GAG 283
DB ALDPKQATNTKAG-PLSIGNSIKRKIN-----GAG 283
QY 133 VORRDAANVAT-----AADQEGQAPSEKSAALSNDVNLIDGKY----- 174
DB VORRDAANVAT-----AADQEGQAPSEKSAALSNDVNLIDGKY----- 174
QY 284 VNKTADVVAQLEAVVKAERITFGDDNSDVKIGLNTLTIKGGAETNALTNNIG 343
DB VNKTADVVAQLEAVVKAERITFGDDNSDVKIGLNTLTIKGGAETNALTNNIG 343
QY 175 YVQADGS-YKKNFALTNGQMLYFSDTGALSSSTSY-----SPSQGTNLVDP-----FSSH 227
DB YVQADGS-YKKNFALTNGQMLYFSDTGALSSSTSY-----SPSQGTNLVDP-----FSSH 227
QY 344 VKEADNSGLKVKLAKTLN---LTVNTTTLNATTTVKVSSSSTTAEILLSLFTQ 400
DB VKEADNSGLKVKLAKTLN---LTVNTTTLNATTTVKVSSSSTTAEILLSLFTQ 400
QY 228 NKAYDSTAKSFELVNGVLTANSWYRPAIGILNCGQWENASNDLRPVLMSWPKDTQYA 287
DB NKAYDSTAKSFELVNGVLTANSWYRPAIGILNCGQWENASNDLRPVLMSWPKDTQYA 287
QY 401 NTGSQSTKTVYGVNGVKFTN-----NAETTAAGTTTRIDKIGFARDGDVDEK 450
DB NTGSQSTKTVYGVNGVKFTN-----NAETTAAGTTTRIDKIGFARDGDVDEK 450
QY 288 VYNNYM-KYLSANETEVNETSQVDLNKEAQSITQIEQKITSDNSTQWLRAMEAFVAA 346
DB VYNNYM-KYLSANETEVNETSQVDLNKEAQSITQIEQKITSDNSTQWLRAMEAFVAA 346
QY 451 QAPYLDKQLKVGSAITDNGIDAGKKISNL-----AKGSSANDA-----VTIEQLKAA 501
DB QAPYLDKQLKVGSAITDNGIDAGKKISNL-----AKGSSANDA-----VTIEQLKAA 501
QY 347 QPKNNSTENFNKGDHLOGGALLYTNSDLTPWANS-----DYRL-LNRTPTQDGTKKY 399
DB QPKNNSTENFNKGDHLOGGALLYTNSDLTPWANS-----DYRL-LNRTPTQDGTKKY 399
QY 502 KPTLN-----AGAGISVTPTEISVDKAGSNVTAPTNYIGVKTLSNDSGTSK 549
DB KPTLN-----AGAGISVTPTEISVDKAGSNVTAPTNYIGVKTLSNDSGTSK 549
QY 400 FTEGGGGYEFLLSNDVNSNPVQAEOL-NQLHYLMNWGDIVM-----GDKDANFD 450
DB FTEGGGGYEFLLSNDVNSNPVQAEOL-NQLHYLMNWGDIVM-----GDKDANFD 450
QY 550 FSVKSGG-----TNNSLVTAHLASYLNEVNRNADTADSLQSFVKEEDDDANAI 598
DB FSVKSGG-----TNNSLVTAHLASYLNEVNRNADTADSLQSFVKEEDDDANAI 598
QY 451 GYRVDAVDNVDNLLQVSNYFKDNKYKVTDSANAL-----AHSILEAWSLNDNQYN 503
DB GYRVDAVDNVDNLLQVSNYFKDNKYKVTDSANAL-----AHSILEAWSLNDNQYN 503
QY 599 TVAKDTTKAGA--VSILKLGKNGLTVAIKDGTTFGLSQDSGLTIGKSLNNDGLTV 656
DB TVAKDTTKAGA--VSILKLGKNGLTVAIKDGTTFGLSQDSGLTIGKSLNNDGLTV 656
QY 504 EDTN-----GTALSDINSSRLT----- 520
DB EDTN-----GTALSDINSSRLT----- 520
QY 657 KDTNEIQVANGIKFTNVNGSPGFGIA--NTARITRDKIGFAGSDGAVDNTKPYLDQ 714
DB KDTNEIQVANGIKFTNVNGSPGFGIA--NTARITRDKIGFAGSDGAVDNTKPYLDQ 714
QY 521 -----1-----SLAVLTQPCQRDLNLSVSNKERRANDTAYG 554
DB -----1-----SLAVLTQPCQRDLNLSVSNKERRANDTAYG 554
QY 715 KLQGVNKTNTNGINAGKAITGLSPTLPSIADQSSRNIELGNTIQD---KOKSNAASIN 771
DB KLQGVNKTNTNGINAGKAITGLSPTLPSIADQSSRNIELGNTIQD---KOKSNAASIN 771

QY 555 DTIPTYSFVRAHDSVQTVIAKIVKBIKIDTNSDGYTFTLDQLDKDAFKIYNEDMAKVNKTY 614
DB DTIPTYSFVRAHDSVQTVIAKIVKBIKIDTNSDGYTFTLDQLDKDAFKIYNEDMAKVNKTY 614
QY 772 DILNTGFNLKNNNNPIDFVSTYDIDVFANGNATTATVTHDTANKTSKVYD--VNVDTT 829
DB DILNTGFNLKNNNNPIDFVSTYDIDVFANGNATTATVTHDTANKTSKVYD--VNVDTT 829
QY 615 THYNIIPAAYALLSNMESVPRVYDLYTDDQGYMAKSPYDAIATMLCGRIAYVSGGQ 674
DB THYNIIPAAYALLSNMESVPRVYDLYTDDQGYMAKSPYDAIATMLCGRIAYVSGGQ 674
QY 830 IHLT-----GTDDNKKLGKVT-----TKLNKTS----- 853
DB IHLT-----GTDDNKKLGKVT-----TKLNKTS----- 853
QY 675 SEEVHKVNGNNQILSSVRYQD--LMSADD-----TQCT-DLSRTSGIAVL 717
DB SEEVHKVNGNNQILSSVRYQD--LMSADD-----TQCT-DLSRTSGIAVL 717
QY 854 -----NGTATNFVNSDEDELVAKDIENLNLAKIHTTGTADTALQTFVVK 906
DB -----NGTATNFVNSDEDELVAKDIENLNLAKIHTTGTADTALQTFVVK 906
QY 718 VSNPDLGDSLTVMGRAHANQAYRPLILGTGQVOSYKDSOTNIIVKYTDANGNL 777
DB VSNPDLGDSLTVMGRAHANQAYRPLILGTGQVOSYKDSOTNIIVKYTDANGNL 777
QY 907 VDENNAD--DANAITVGQKANNQVNTLTKEGNGL-----NI--KTDKNGT 952
DB VDENNAD--DANAITVGQKANNQVNTLTKEGNGL-----NI--KTDKNGT 952
QY 778 F---TADDIK-GYSTVDMGYLAIVVPVGAQD---QD-VRVAADTNQKADGSLKTSAA 829
DB F---TADDIK-GYSTVDMGYLAIVVPVGAQD---QD-VRVAADTNQKADGSLKTSAA 829
QY 953 FGINTTSGLKAGKSTLN-DGGLSIKNPSTGSEQIVGADGVKFAKVNNGVVGAGIDGTT 1011
DB FGINTTSGLKAGKSTLN-DGGLSIKNPSTGSEQIVGADGVKFAKVNNGVVGAGIDGTT 1011
QY 830 LDSQVIYEGFSNFQDPANNDADVTNKKIAE--NADFFKKLGITSEMAPQYVSATDGSF 886
DB LDSQVIYEGFSNFQDPANNDADVTNKKIAE--NADFFKKLGITSEMAPQYVSATDGSF 886
QY 1012 ITRDEI--GFTG-----TNGSLDKSKPHLSKDGINAGGKKTNIQSGEIAQNSHDAVTGK 1065
DB ITRDEI--GFTG-----TNGSLDKSKPHLSKDGINAGGKKTNIQSGEIAQNSHDAVTGK 1065
QY 887 LDSIIQNGYAFSDRYDLAMSKNNKYGSKDDLAKALKALH-----ANGTQAI 932
DB LDSIIQNGYAFSDRYDLAMSKNNKYGSKDDLAKALKALH-----ANGTQAI 932
QY 1066 I-----YDLKTELENKISS--TAKTAQNSLHEFSVADQGNFTVSNPYSSY 1110
DB I-----YDLKTELENKISS--TAKTAQNSLHEFSVADQGNFTVSNPYSSY 1110
QY 933 ADWVPDQIYOLPGEVVTAK-----RTNSYGNPTFPDAYINNA----- 969
DB ADWVPDQIYOLPGEVVTAK-----RTNSYGNPTFPDAYINNA----- 969
QY 1111 DTSKTSVDITFAGENGITTKVNGVVRVIGDITGKLTTPKLTGVNNGNGKVIDSONGQ 1170
DB DTSKTSVDITFAGENGITTKVNGVVRVIGDITGKLTTPKLTGVNNGNGKVIDSONGQ 1170
QY 970 -----LVATNTKSSGSDYQAYGGAFLDELKAKYPMFTVMNISTGKPIDSTIKQW 1022
DB -----LVATNTKSSGSDYQAYGGAFLDELKAKYPMFTVMNISTGKPIDSTIKQW 1022
QY 1171 TITGLSNTLIANTNDKGSVRTTEQGNIIKDEKTRAASI--VDVLSA----- 1215
DB TITGLSNTLIANTNDKGSVRTTEQGNIIKDEKTRAASI--VDVLSA----- 1215
QY 1023 EAKYFNGTNVLGKAGYVLSDDATGKYFTVN--ENGDFLPASFTGQNAKTG--FYD-- 1076
DB EAKYFNGTNVLGKAGYVLSDDATGKYFTVN--ENGDFLPASFTGQNAKTG--FYD-- 1076
QY 1216 -----GFNLQNGEAV---DFVSTYDVTNFDAGNATTAKVTYDDTSKTSKVYDVNVD 1265
DB -----GFNLQNGEAV---DFVSTYDVTNFDAGNATTAKVTYDDTSKTSKVYDVNVD 1265
QY 1077 -----GTGMAYYSTSGNAKVNFSIYEG-----GHYYPFDKGHNV 1111
DB -----GTGMAYYSTSGNAKVNFSIYEG-----GHYYPFDKGHNV 1111
QY 1266 DTTIEVKDKKLGKVTTLTSTGTGANKFALS--NQATGDALVKASDIVAHLNLTSLGDIQTA 1324
DB DTTIEVKDKKLGKVTTLTSTGTGANKFALS--NQATGDALVKASDIVAHLNLTSLGDIQTA 1324
QY 1112 TGSYKAECDNDYFLPNGLQMEDAIYQDAQGSYYGRTGILY---KGNWNPFPVDPNNA 1168
DB TGSYKAECDNDYFLPNGLQMEDAIYQDAQGSYYGRTGILY---KGNWNPFPVDPNNA 1168
QY 1325 KGASQANN-----SAGYVDADGNKVIYDSTDNKYQAKNDGTVDKTKEVAK 1370
DB KGASQANN-----SAGYVDADGNKVIYDSTDNKYQAKNDGTVDKTKEVAK 1370
QY 1169 NKTVPFRYFDANNVMA--IGYRNMYGQTYYPEDENGFOAKGQLLTDD---KGTHYFDEDNGA 1223
DB NKTVPFRYFDANNVMA--IGYRNMYGQTYYPEDENGFOAKGQLLTDD---KGTHYFDEDNGA 1223
QY 1371 DKLVQAQTPDGTLAQMNKVSINKEQVNDAN---KQGINEDNAFVKGLEKASDN-- 1424
DB DKLVQAQTPDGTLAQMNKVSINKEQVNDAN---KQGINEDNAFVKGLEKASDN-- 1424
QY 1224 MAKNFVNVGDDWYWDGNGNAVKGQYPVNNOILYPNPETGVQVK--GQFITDAQGRITSY 1281
DB MAKNFVNVGDDWYWDGNGNAVKGQYPVNNOILYPNPETGVQVK--GQFITDAQGRITSY 1281
QY 1425 KTKNAAVTVGD-----LNAV-AQTP-----LTFAGDTGTTAKKLGELTIIKGGQT-- 1468
DB KTKNAAVTVGD-----LNAV-AQTP-----LTFAGDTGTTAKKLGELTIIKGGQT-- 1468
QY 1282 YDANSALKSSGFFTPNGSDWYEAENGYYKGFQVAENQDQWYFDTTGTQAKAAKV 1341
DB YDANSALKSSGFFTPNGSDWYEAENGYYKGFQVAENQDQWYFDTTGTQAKAAKV 1341
QY 1469 -DTNK-----LTDNNIGVVAGTDGFTVVKLAKOLT-----NLNSVNAAGTKI 1508
DB -DTNK-----LTDNNIGVVAGTDGFTVVKLAKOLT-----NLNSVNAAGTKI 1508
QY 1342 DGRDLYFNPDSGVQVKGDFATDESNTSFYHG---DNGDKVVGFFFTGNNAWYADNNG 1398
DB DGRDLYFNPDSGVQVKGDFATDESNTSFYHG---DNGDKVVGFFFTGNNAWYADNNG 1398
QY 1509 DDGKVSFVDSGG-QAK-----ANTPVLNGLDLGKVI-----S 1542
DB DDGKVSFVDSGG-QAK-----ANTPVLNGLDLGKVI-----S 1542
QY 1399 NLVKGQFQELD-KWYHFDVETVQQAAGAA-----LVNGQOLYFVDVDSGIO-----VK 1444
DB NLVKGQFQELD-KWYHFDVETVQQAAGAA-----LVNGQOLYFVDVDSGIO-----VK 1444
QY 1543 NVYKGTQKTDAAVQQLNEVRNLLGLNAGNADGNQVNIADIKDPNSGSSSNRTVIK 1602
DB NVYKGTQKTDAAVQQLNEVRNLLGLNAGNADGNQVNIADIKDPNSGSSSNRTVIK 1602
QY 1445 GDFVTDGQNTSYDVNSG-----DKKNGFTTGD--NAN-----YYA 1481
DB GDFVTDGQNTSYDVNSG-----DKKNGFTTGD--NAN-----YYA 1481
QY 1603 ACTVLGGKGNNDTEKLTATGGIQVGVDKGN---ANGDLNWNVWTKQKSGKALLATYNA 1659
DB ACTVLGGKGNNDTEKLTATGGIQVGVDKGN---ANGDLNWNVWTKQKSGKALLATYNA 1659
QY 1482 DQ-----GNLAKGRKSIDNQL--YFDPATGQVKGQLVSIIDGRNYYFDSGSGN----- 1529
DB DQ-----GNLAKGRKSIDNQL--YFDPATGQVKGQLVSIIDGRNYYFDSGSGN----- 1529

Db 1660 AGQTNVLTNNPAAIRDRNEQGRFFHVNDGNEP-----VQGRNGIDSSASGKHSVAIG 1715
QY 1530 -----MAKNRFVRIGDQWYFNGDGAAT 1552
Db 1716 FOAKADGEAAVAIGRQTQAGNQSIGAIGDNAQT 1748
RESULT 5
US-11-013-759-4
; Sequence 4, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN OF MORAXELLA
; FILE REFERENCE: 1038-921MIS:jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2047
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-4
Query Match 4.4%; Score 363; DB 7; Length 2047;
Best Local Similarity 20.6%; Pred. No. 2.6e-11;
Matches 377; Conservative 196; Mismatches 682; Indels 578; Gaps 86;
QY 18 TIAVASIGLSVLGAGTVSAEDKVANDTTAAQATVGVDTCGQOATTNDANTN-----TTDTD 73
Db 251 TIAIGSDATSSSGAIALGA-----GTRAOLGSGALGQGSVVTSQDNNRSPAYTPNTQ 304
QY 74 TADQSAN-TNQDQAGSDQSNQDQAKQDTANTDRNQADNSQTDNNQATQATSPATDGTG 132
Db 305 ALDPKQATNNTKAG-PLSIGSNSIKRKIINV-----GAG 338
QY 133 VQRDRANVAT-----AADQEGQAPSEQKSAALSDNVKLDGKY----- 174
Db 339 VNKTDVAVNAQLEAVVAKKERRITFGDDNDSDVKIGLDNTLTIRKGAETNALTDNNIG 398
QY 175 YYVQADGS-YKKNFALTNGQMLYFSDTGALGSTSTY---SFSQGTNLVDD---PSSH 227
Db 399 VVKEADNSGLVKVLAKTLNN---LTEVNTITTLNATTIVKVGSSSTTAEILLSLSLTFTQP 455
QY 228 NKAYDSTAKSFELVNGYLTANSYRPAIGILRNQGTWEASNENDLRPVLMSWPDKDTQVA 287
Db 456 NTCQSSTKTVGVGVKFTN-----NAETTAAGTTRITRDKIGFARDGVDDEK 505
QY 288 YVNMN-KYLSANETEVNTEQSDLNKEAQSIQTKEIKITSDNSTQMLRTAMEAFVAA 346
Db 506 QAPYLDKKQLKVGSAITIDNGIDAGNKKISNL-----AKGSSANDA---VTIEQLKAA 556
QY 347 QPKWNMSTENFNKGDHLOGGALLYTNSDLTPWANS-----DYRL-LNRPTQODGTCKY 399
Db 557 KPTLN-----AGAGISVTPTEISVDKSGNVNAPTNYIGVKTITELNSDGTSDK 604
QY 400 FTEGGGEGYFLLSNDVNSNPVVQAEQL-NQLHYLMNWGDIVW-----GDKDANFD 450
Db 605 FSVKGS-----TNSLVTAHLASYLNEVNRADTSALQSFVTKVEDDDANAI 653
QY 451 GVRVADVNNADLLQVYSYFKDNYKVTDSEANAL-----AHISILEAWSLNDNQVN 503
Db 654 TVAKDITTKNAGA--VSILKLKGNGLTVATVKDQGTVTFGLSQDSGLTIGKSTLNDGLTV 711
QY 504 EDTN-----GTALSIDNSSRLT----- 520

Db 712 KOTNEQIQVGANGIKFTNVGNSNPGTGIA--NTAITRDKIGFAGSDGAVDINKPYLDQD 769
QY 521 -----SLAVLTQPGQRIIDLNLNLSVESVKNKERANDTAYG 554
Db 770 KLVQGVNKITNTGINAGGKAITGLSPTLPSIAIQSSRNIELGNTIQD--KDKSNAASIN 826
QY 555 DTIPYISFVRAHDSVQTVIAKIVKEKIDTNSDGYTFTLDQLKDAFKIYNEDMAKVNKY 614
Db 827 DILNTGFNLKNNNPIDFVSTYDIDVFANGNATTAATVTHDTANKTSKVYVD--VNVDDTT 884
QY 615 THYNIPAAVALLSNMESVPRVYVYDLYTDDGOYMAKSPYYDAIATMLQGRYAYVSGGQ 674
Db 885 IHLT-----GTDDNKKLGKVT-----TKLNKTSA----- 908
QY 675 SEEVHKVNGNNOILSVRYGQD--LMSADD-----TQGT-DLRSRTSLGLVTL 717
Db 909 -----NGTATNFNVNSSDEDAVLNAKIDIAENLNTLAKEIHHTTKGTADTALQTFVVK 961
QY 718 VSNDPNLDLGGDSLTVNMGRAHANQAYRPLILGTGQVOSYLKSDSTNIVKYTDANGNLT 777
Db 962 VDENNNAD--DANAITVGQKNNANQVNTLTLLKGENL-----NI--KTDKNGTVT 1007
QY 778 F---TADDIK-GYSTVDMSGYLAVVPVGAKD---QGD--VRVAADTNQKADGKSLKTSAA 829
Db 1008 FGINTTSGLKAKGSTLN-DGGLSIKNPTGSEQIQVGADGVKFAKVNNGVVGAGIDGTTT 1066
QY 830 LDSQVIEGFSNFPQDPAANDADYTNKKIAB--NADFFKKLGITFEMAPQVYSATDGSF 886
Db 1067 ITRDEI--GFTG---TNGSLDKSPHLSKDGINAGGKKITNIQSIEIAQNSHDAVTGGK 1120
QY 887 LDSIIQNGYAFSDRYDLAMSKNNKYGSKDDLANALKALH-----ANGIOAI 932
Db 1121 I-----YDLKTELENKISS--TAKTAQNSLHEFSVADEQGNFTVSNPYSS 1165
QY 933 ADWVPDQIYQLPGEVVTAK-----RTNSYGNPTFDAYINNA----- 969
Db 1166 DISKTSVDITFAGENGITTKVNGKVVVRGIDQTKGLTTPKLVGNNNGKGIIVDSQGN 1225
QY 970 -----LYATNTKSGSDYQAYQGGAFDELAKAYPDMFTVMNISTGKPIDPSTKIKOW 1022
Db 1226 TITGLSNTLANVTNDKGSVRTTEQGNIIKDEKTRAASI--VDVLSA----- 1270
QY 1023 EAKYFNGTIVLKGAGYVLSDDATGKYFTVN--ENGDFLPASFTGDNQAKTG-FYYD--- 1076
Db 1271 -----GFNLQNGEAV---DFVSTYDTVNFADGNATTAKVTYDDTSTKSVVYDVNVD 1320
QY 1077 -----GTGMAYYSTSGNKAVNSFIYEG---GHYYFPDKDGHMV 1111
Db 1321 DTTIEVKDKKLGKVTTLTSTGTGANKFALS--NQATGDALVKASDIVAHLNTLSGDIQTA 1379
QY 1112 TGSYKAEDGNDYFPLPNGIQMRDAIYQDAQNSYYYGRTGILY---KGNWYFPFVDPNNA 1168
Db 1380 KGASQANN-----SAGYVDADGNKVIYDSTDNKYQAKNDGTVDKTEKAV 1425
QY 1169 NKTVPFYPDANNVMA-IGYRNMYGQTYYPEDENGFOAKQLLTTD----KGTHYFDEDNKA 1223
Db 1426 DXLVAQAQPTPDGTTLAQMNVKSVINKEQVNDAN---KKGINEDNFAVKGLEKASADN-- 1479
QY 1224 MAKNFVNVYDQWYMDGNGNAVKQYPVNNQILYFNPETGVQVK--GQFITDAQRTSY 1281
Db 1480 KTNAAVTVGD-----LNAV-AQTP-----LTFAGDTGTTAKKLGETLTIKGGQT-- 1523
QY 1282 YDANGALKSSGFFTPNGSDWYEAENGYYVYKGFQVAENQDQWYDPDTGTGQAKAAKV 1341
Db 1524 -DTNK-----LTDNNIGVVAGTDGFTVKLAKDLT-----NLNSVNAAGTKI 1563
QY 1342 DGRDLYFPNPDGSGVQVKGDFATDESNGTSFYHG---DNGKVVGGFTTGNNAWYADNNG 1398
Db 1564 DDKGSFVDSG--QAK-----ANTPVLNANGLDLGGKVI-----S 1597
QY 1399 NLVKGFQEID-GKWYHFDEVTGQAKGAA-----LVNQQLYFVDVDSGIO-----VK 1444

Db 1598 NVGKGTDTDAANVQQLNEVRNLLGLGNAGNDNADGNQVNIADIKKDPNSGSSSRTVIK 1657

Qy 1445 GDFVTDQSGNTSYDVNSG-----DKVNGFFTTGD--NAW-----YYA 1481

Db 1658 AGTVLGGKGNDEKLGATGGIQQVDRKDN---ANGDLNVVWVKTKQDKGSKALLATYNA 1714

Qy 1482 DGO-----GNLAKGRKSIDNQDL-YFDPATGKQVKQLVSDIGRNYFYDFSGSN----- 1529

Db 1715 AGQNTYLTNNPAEADIRNEQGRFFHFVNDGNQEP-----VVOGRNGIDSSASGKSHVAIG 1770

Qy 1530 -----MAKNRFVRIGDQWIFGNDGAAT 1552

Db 1771 FOAKADGEAAVAIGRQTAQGNQSIAGDNAQAT 1803

RESULT 6

US-11-013-759-7

; Sequence 7, Application US/11013759

; Publication No. US20050249747A1

; GENERAL INFORMATION:

; APPLICANT: Loomis, Sheena M.

; APPLICANT: Sasaki, Ken

; APPLICANT: Yang, Yan Ping

; APPLICANT: Klein, Michel H.

; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE

; TITLE OF INVENTION: PROTEIN OF MORAXELLA

; FILE REFERENCE: 1038-921MIS:jb

; CURRENT APPLICATION NUMBER: US/11/013,759

; CURRENT FILING DATE: 2004-12-16

; PRIOR APPLICATION NUMBER: US/09/361,619

; PRIOR FILING DATE: 1999-07-27

; NUMBER OF SEQ ID NOS: 32

; SOFTWARE: Patent In Ver. 2.1

; SEQ ID NO 7

; LENGTH: 2047

; TYPE: PRT

; ORGANISM: Moraxella catarrhalis

US-11-013-759-7

Query Match 4.4%; Score 363; DB 7; Length 2047;

Best Local Similarity 20.6%; Pred. No. 2.6e-11;

Matches 377; Conservative 196; Mismatches 682; Indels 578; Gaps 86;

Qy 18 TIIVASIGLVSLVAGTVAEDKAVNDTTAAQTAVGVDTGQDOATTNDANTN-----TTDTD 73

Db 251 TIAIGSDATSSSLGATLGA-----GTRALQGSIALGGSVVVTQSDNNSRPAYTPNQ 304

Qy 74 TADQSAN-TNQDQAGSDQSNQDOAKQDTANTDRNOADNSQTNNOATQOATSPTDGT 132

Db 305 ALDPKFOATNTKAG-PLSIGNSIKRKIINV-----GAG 338

Qy 133 VORRDAANVAT-----AADQSGQAPSEQKSAALSIDNVKLIIDGKY----- 174

Db 339 VNKTDAVVAQLEAVVWAKERRITQGGDNSTVDVKIGLDNTLITKGGAEETALTDNNIG 398

Qy 175 YVVOADGS-YKKNFAITVNGQMLYFSDTCALSSSTSY-----SFSQGTINLVDD---FSSH 227

Db 399 VVKEADNSGLVKLAKTLN---LTVNTTLNATTVKVGSSSTTAEILSDSLFTQ 455

Qy 228 NKAYDSTAKSFELVNGYLFWNSWYRPAIGLRNQTWAEASNENDLRPVLMSWMPDKDTQVA 287

Db 456 NTGSQSTKTVYGVNGVKFTN-----NAETAAIGTTRITRDKIGFARDGVDDEK 505

Qy 288 YVNYMN-KYLSANEIEVTNETSQVDLANKAQSLOTKIEOKITSDNSTQWLRTAMEAFVAA 346

Db 506 QAPYLDKKQLKVGVSVAITIDNGIDAGNKKISNL-----AKGSSANDA-----VTIEQLKAA 556

Qy 347 QPKWNNSTENFNKGDLQGGALLYTNSDLTPWANS-----DYRL-LNRTPQDDTKKY 399

Db 557 KPTLN-----ACAGISVTPTEISVDAKSGNVTAPYNTGVKVTIELNSDGTSDK 604

Qy 400 FTEGEGGYEFLLSNDVDNSNPVQAEQL-NOLHYLMNWGDIVM-----GDKDANFD 450

Db 605 FSVKSGS-----TNNSLVTAHFLASLYLNEVNRRTADSAQLQSFTVKEDEDDDDANAI 653

Qy 451 GVRVDAVDNVADLLQVSYSNYFKDNYKVYTDSEANAL-----AHISILEAMSLNDNQYN 503

Db 654 TVAKDTTKRAGA--VSILKLGKNGGLTVATKDGTVTFGLSQDSGLTIGKSTLNDGLTV 711

Qy 504 EDYN-----GTALSIDSRLT----- 520

Db 712 KDTNEQIQVGANGIKFTNVNGSNPGTIGIA--NTARITRDKIGFAGSDGAVDTNPKPYLDQD 769

Qy 521 -----SLAVLTQPGQORIDLSNLSISVNVKERANDTAYG 554

Db 770 KLQGVNWKITNTGINAGKAITGLSPISIADQSSRNIELGNTIQD-----KDKSNAASIN 826

Qy 555 DTIPTYSFVRAHDSVQTVIAKIVKEKIDTNSDGYTFTLDQLDKAFKIVNEDMAKVNKTY 614

Db 827 DILNTGFNLKNNNPIDFVSTYDIVDFANGNATTAIVTHTTANKTSKVYD--VWUDDTT 884

Qy 615 THYNIPAAVALLSNMESVPRVYDLYTDDGQYMAKSPYDAIATMLQGRYAYVSGGQ 674

Db 885 IHLT-----GTDDNKKLGVKT-----TKLNKTS----- 908

Qy 675 SEEVHKVNGNNQILSSVRYQD--LMSADD-----TOGT-DLSTSGLVTL 717

Db 909 -----NGTATNFNVNSSDEDALYNKADIAENLNTLAKEIHTTKGTADTALQTFVVK 961

Qy 718 VSNPDLGLGDSLTVMNGRAHQAQVRLILGTQGVQSYLQSDTNI VKYTDANGNLT 777

Db 962 VENNAD--DANAITVGKXANNQVNTLTLXENGL-----NI--KTDKNGT 1007

Qy 778 P---TADDIK-GYSTVDMSGYLAVVYPVGAQD---QD-VRVAADTNQKADGKSLKTSAA 829

Db 1008 FGINTTSGLKAGKSTLN-DGGLSIKNPTGSEQIQVGADGVKPAKVNNGVVGAGIDGTT 1066

Qy 830 LDSQVIYEGFSNFPQDPANNDADYTNKXIAB---NADFFKLGLTSEMAPQYVSATDGSF 886

Db 1067 ITRDEI--GFTG---TNGSLDKSPHLSKOGINAGGKITTNTQSGEIAQNSHDAVTGCK 1120

Qy 887 LDSIIQNGYAFSDRYDLAMSNNKYGSKDDLALAKALH-----ANGLOAI 932

Db 1121 I-----YDLKTELENKISS--TAKTAQNSLHEFSVADQGNFTVSNPYSSY 1165

Qy 933 ADMVPDQIYQLFEEVVTAK-----RTNSYGNPTFDAYINNA----- 969

Db 1166 DTSKTSVITFAGENGITTKVNGVVRVIGIDQTKGLTTPKLTGVNNGKGIIVDSQNGQ 1225

Qy 970 -----LYATWTKSSGSYQAYGGAFLDELKAKYPDMPTVMNISTGKPIDPSTKIKOW 1022

Db 1226 TITGLSNTLANVTNDKGSVRTTEQGNIIKDEDKTRAASI--VDVLSA----- 1270

Qy 1023 EAKYFNGTNVLGKAGYVLSDDATGKYFTVN--ENGDFLPASFTGQNAKTG-FYYP----- 1076

Db 1271 -----GFLQNGEAV---DFVSTYDTVNFADGNATTAKVYDTSKTSKVYVDNV 1320

Qy 1077 -----GTGMAYYSTGSKAVNSPIYEG-----GHVYFPDKGHMV 1111

Db 1321 DTTIEVKDKLGVKTKTLTSTGTGANKFALS--NQATGDALVKASDIVAHLNTLSGDIQTA 1379

Qy 1112 TGSYKAEDGNDYFPLPNGIQMRDAIYQDAQGSYYYGRTGILY---KGDNWPFPVDPNNA 1168

Db 1380 KGASQANN-----SAGYVDADGNKVIYDSTDNKYQAKNDGTVDKTKREVAK 1425

Qy 1169 NKTVPFYPDANNVMA-IGYRNMVGYTYFDENGFOAKQLLTTD-----KGTHYFDEDNCA 1223

Db 1426 DKLVAQAQTPDGTTLAQMNVKSVINKEQVNDAN---KKGINEDNAFVKLEKASDN-- 1479

Qy 1224 MAKNFVNVGDDWYMDGNGNAVKGOYPVNNQILYFNPETGVQVK--GOFITDAQORTSY 1281

Db 1480 KTKNAVTVGD-----LNAV-AQTP-----LTFAGDTGTAKKLGTLTIKGGQT-- 1523

Qy 1282 YDANSALKSGSFFTPNGSDMYAENGYYVYKGFQVAENQDQWYFPDQTTGKQAKGAAKV 1341

Db 1524 -DTNK-----LTDNNIGVVAGTDGFTVKLAKDLT-----NLNSVNAAGGTGI 1563

Qy	1239	MDGNGNAVKGOYPVNNQILLYENPBTGVQVK---GQFIITDAQRTSYYPDANGSALKSGSFFT	12996
Db	1500	-----LNAV-AQTP-----LTFAGTGTAKKLGETLTIKGGQT---DTNK-----LT	1538
Qy	1297	PNGSDWYYAENGYYGVFKOVAENQDQWYIFDQTTGGKQAAKVDGRDLYENPDSGVQV	1356
Db	1539	DNNTGVVAGTGDFTVKLAKDLT-----NLNSVNAAGTKIDEGKISF-VDANGQA	1586
Qy	1357	KGDPATDESNTSFYHG---DNGDKVYVGGFFTTGNNAWYYADNNGNLVKGFQEBID-GKWY	1412
Db	1587	K-----ANTPVLNANGLDLGGKVI-----SNVGKGTKDTDAANVQ	1621
Qy	1413	HFDEVTCQQAAGALVNGQOLYF-----DVDSGIQ-----VKGFVTFDGGNTSYVDVNS	1462
Db	1622	QLNEVRNLLGLGNDNADGNQVNTADIKKDPNSGSSSNRTVIKAGTVLGGKGNNDTEKLAT	1681
Qy	1463	G-----DKKXNGFFTTGD-NAW-----YYADQO-----GNLAKGKESI	1494
Db	1582	GGVQGVVDKQGN-----ANGDLSNVWVTQKQSGKALLATYNAAGQTYVTNNPAAIDRI	1738
Qy	1495	DNQDL-YFDPATGKQVKQLVSDIGRNYYPFDSGSN-----MAKNRFVR	1537
Db	1739	NEQIRFHVNDGNQEP-----VVQGRNGIDSSASGKHSVAIGFQAKADGAAVAIGRQTQ	1794
Qy	1538	IGDQWIFYGNDGAAT	1552
Db	1795	AGNOSIAIGDNAOAT	1809

RESULT 8

US-11-013-759-11

; Sequence 11, Application US/11013759

; Publication No. US20050249747A1

GENERAL INFORMATION:

APPLICANT: Loosmore, Sheena M.

; APPLICANT: Sasaki, Ken

; APPLICANT: Yang, Yan Ping

; APPLICANT: Klein, Michel H.

; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE

; TITLE OF INVENTION: PROTEIN OF MORAXELLA

FILE REFERENCE: 1038-921MIS

; CURRENT APPLICATION NUMBER: US/11/013,759
 ; CURRENT PCT NUMBER: 2004/00015

; CURRENT FILING DATE: 2004-12-16

;; PRIOR APPLICATION NUMBER: US/09/361,619
;; PRIOR FILING DATE: 1998-07-27

;; PRIOR FILING DATE: 1999-07-27
NUMBER OF SEQ ID NOS: 30

; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: DataIn Vm

```

; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 11

```

```

; SEQ ID NO 11
; LENGTH: 2314

```

; LENGTH: 2314
: TYPE: DPT

1. LIFE: FRI
2. ORGANISM: MO

ORGANISM: *MOTAXELLA CATALINENSIS*
US-11-013-759-11

TT-CCA-CTA-TT-CG

Query Match 3.7%: Score 306.5: DB 7: Length 2314:

Best Local Similarity 21.0%: Pred. No. 2.6e-08:

Matches	353:	Conservative	205:	Mismatches	637:	Indels	489:	Gaps	91:
---------	------	--------------	------	------------	------	--------	------	------	-----

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

QY 23 SIGLSLVGAGTVAEDKV-ANDTTAQAT-----VGVD-----TGOD---- 58

[illegible]

Db 776 SVGFVS-----TYNTVDFIDGNATTAKVTYDETNQTSKVTYDVNVDEKTIELTGDNGKTN 830

.....

QY 59 - - - - - QATTDANTNTTDTADQSANTNQDAGSDQSNNQDQAKQDTANTDRNQADN 111

[illegible]

Db 831 KIGVKTTLTTNANGKATNFSTDNDALVN---AKDIAENLTLAKE-IHTTKGTADT 885

Qy	222	DD	----	PSHKNKAVDST	-----	AKSFELVNGYLTANSWTRPAGILRNGQTMWASNE	268									
Db	1000	KDOIGFTCANGSLD	TTKPHLT	KDKLKVGEVEIT	TNTGINAGG--	KKIITNIQSGDITQNSND	1057									
Qy	269	-----	NDLRPVLMSW	PKDQVAVV	-NYNMYKLSANET	-----	EVTNETSQVDLNKE	315								
Db	1058	AVTGRVYDUKTELES	-----	KINSAAKTAQNSLH	FEFSADEQGNHFTVSNPYSYDTSKT		1113									
Qy	316	AQSIQTQKIEQKITS	SDNSTOWL	RAMEAFVAAQPKWNN	STENFNKGDHLQCGALLYTNSD-	374										
Db	1114	SDVITPAGENGII	TKUNKGV	VRVID	-----	QYK-GLTTPKLT	VGNV-NKGKIVIDSKDG	1166								
Qy	375	---	LTPWANSDYRL	LNPTQODG	TKKYFTTEGEGGYEF	LNSDNVNSNPVVOAEQLNQL	431									
Db	1167	QNTITGLSNT	---	LANTV	---	NDGAGHALSQG	-----	LANDTDKTRAASIGDVLNAG	1212							
Qy	432	HYLMWNGDI	VMGDKAN	FGDVR	VDADVNVNADLLQVY	SNYPKDNKYKTDSEANALAHISI	491									
Db	1213	FNLOGNGEAV-	-DFVSTYD	--	TVDFFDG-NATTAKV	-----	TYDDTSKTSKVYVDVNV	1260								
Qy	492	LEAWSLNDN	QYNEDTNG	TALSI	DNSSRLTSLAVLT	KPGQQRIDLSNLISBSVNKERANDT	551									
Db	1261	-----	DNKTIETVSD	KKLGVKTT	-----	LTKISA	-----	NGNATKFSAAD-	1296							
Qy	552	AYGDTIPTYS	FSVRAH	-----	DSEVQTVIAKIV	KEBKIDT	-----	NSDGYFTFLDQLDKAFKIY	603							
Db	1297	--	GDALVKASDI	ATHLNTLAG	DIQT--	AKGASQASSAS	VVDADGNKVIYDSTD	KKYQV	1352							
Qy	604	NEDMAKUNKY	TYHYNIP	PAAYALL	LSNMESVPRVY	GDLYTDDQYQWAKSPYDATA	TML	663								
Db	1353	N-DKGQVDK	-----	NKEVAKDKL	VAAQATPDG	-----	TLA	1381								
Qy	664	QGR1-AVSGGQ	SEEVHKVNG	NNQILLSSVRY	GDLMASADDTQCTG	DLTSRSGLTVLYSNDP	722									
Db	1382	QMNKSVIN	KEQVNDANK	KQINEDNA	FIKGLN--	AAKDTYKNA	AVTVGDLNVAO	ATP	1439							
Qy	723	NLDLG	-----	GDSLTVNMG	RAHANOAYR	---	PLILGT	KDGVOSYILK	SDTNI	VKYT	770					
Db	1440	LTPAGDTGT	TAKKLG	ETLITIKG	QTDNTKLT	DNNIGWAGT	-DGFTVK	LAKDLTNL	-NSV	1497						
Qy	771	DANGLT	FTADDIK	GYSTVDM	SGYLAVWPV	GAKDQDV	VAAD	TNQKADG	SLKT-SAA	829						
Db	1498	NAGG	-----	TRIDBKGI	SFVDANG	QAKANTFVLS	ANGLDL	-----	GGKRIS	WIGAA	1543					
Qy	830	LDSQVIEG	FSNFQD	FANNAD	YTNKKIA	ENADFPKKLG	LIGITSP	EMAPQV	VSATD	SGFLDS	889					
Db	1544	VDD	-----	NDAVNF	KQF--	NEVAKTVN	LNQNSG	ASLFP	VTDANG	KPI	NGT	DGP	KQA	1597		
Qy	890	I--	IQNGYAF	SDRYDL	AMS	KNK-K	-YGS	KODLAN	AL	KALHANG	IQAI	ADW	VPDQ	IYOL	PGE	946
Db	1598	IKGADG	KYHANANG	VPVDK	GKPI	TDADK	LJANL--	AAHG	KPLDA	-----	GH	1642				
Qy	947	EVTAKRT	WSYGNPT	FDAYIN	ALYATNTKS	-----	SGSDYQA	984								
Db	1643	QVVASLG	SGNS	-----	DAITLT	NTIKSTLP	QIDTPT	GNANAG	QASLPS	LAAQ	1691					
Qy	985	QYGAFL	DELK	KAKY	PDWFTV	NMISTG	KPIDP	SPKIKQ	WEAKYFNG	NVLG	KGAGY	VYL	SDD	1044		
Db																

Db 1872 DTDAAVQQLNEVRNLLGLGN-----AGNDNA-----DGNQVN---TA 1906
QY 1274 DAQGRSYSDANSALKSSGFFTPNGSDWYEAENGYYVYKGFQVAENQDQWYFDQTTGK 1333
Db 1907 DIK-----KDPNSG-----SSNRTVIKAGTVLGG-----KGND-----TEK 1939
QY 1334 QAKGAADVDRDLFYNPDSGVQVKGDFATDESNTSFYHGDNGD-----KV 1379
Db 1940 LATG-----GVQV-----GVQKQDNA-----NGDLSNVWVKTKQKGSKEA 1974
QY 1380 VGGFFTTGNNAWYADNNGNLVKFQEIIDGKWHFDEVYTGQQAAGALVNGQQLYFDVDS 1439
Db 1975 LLATYNAAGQTNLYTNPAEADRINEQGIRFHVND--GNQ---BPVVOGRN---GIDS 2026
QY 1440 -----4-----GIQVKGDFVTDGCGNTSYDVNSGDKKVGFFTTGDNAMYAD-----GQ 1484
Db 2027 SASGKHSVALGFAK-----ADGEAAVA---IGRQTAGNQSTAIGDNAQATGDSIAIGT 2079
QY 1485 GNLAKGRKSIDNQDLYFDPATGKQVKGQLVSDGRNYYFDS-----GSGN---MAKNRF 1535
Db 2080 GNVVTGKHS-----GAIGDPSTVK--ADNSYSVGNNGHOFIDATQTDVFGVGNNTVTSES 2133
QY 1536 VRIG 1539
Db 2134 VALG 2137
RESULT 9
US-11-022-562-228
; Sequence 228, Application US/11022562
; Publication No. US20050249742A1
; GENERAL INFORMATION:
; APPLICANT: Shisong, Jiang M.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING
; FILE REFERENCE: A CYTOTOXIC T LYMPHOCYTE IMMUNE RESPONSE
; CURRENT FILING DATE: 2004-12-22
; PRIOR APPLICATION NUMBER: PCT/US03/20322
; PRIOR FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: 60/392718
; PRIOR FILING DATE: 2002-06-27
; NUMBER OF SEQ ID NOS: 340
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 228
; LENGTH: 396
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-11-022-562-228
Query Match 3.7%; Score 302.5; DB 7; Length 396;
Best Local Similarity 24.8%; Pred. No. 4.1e-09;
Matches 108; Conservative 67; Mismatches 176; Indels 85; Gaps 18;
QY 1121 NDYY-----FLPNGIQMRDAIYQDAQNSYYIGRTGILYKGDNWPFPDNNANKTVFRY 1175
Db 1 NEYYPEIIVLNPFTFHKKYINILDSSEFYKWSLEG-----SDFILVRY 44
QY 1176 FDANN---VMAIGYRNMYGOTYYFDENGFOAK-----GQLTDDKGYHFDENGAMA 1225
Db 45 LEBSNKKILQKIRIKILSNTKSFNMSIDFKDIKLSLGYIMSNFKS---FNSNELDR 101
QY 1226 KN--KFNVNGDDWYDMGNGNAVGQYPVNNQILYFNPEYVQVKGQFIDTAQ---GRTSY 1281
Db 102 DHLGFKIINDKTYTYDEASKLVKGLININLSLFFDP-----IESNLVTGWQTIINGKKYY 156
QY 1282 YDANSALKSSGFFTPNGSDWYEAENGYYVYKGFQVAENQDQWYF--DQTTGKQAKGAA 1339
Db 157 FDINTGA-ASTSYKIINGKHFFYNNGNVMQLG---VFKGPDGPEYFAPANTONNIEGQA 212
QY 1340 KVDGRDLFYNPDSGVQVKGDFATDESNTSFYHGDNGDKVVGFFTTGNNAWYADNNGN 1399

Db 213 -----IVYQSKFLT---LNGKKYFNDNSKAVTGWQTIIDGKKYFNLNTAE 255
QY 1400 LVKQFEIDGKWHFDEVYTGQQAAGALVNGQQLYFDVDSGIQVRKGFVTDGQNTSY- 1458
Db 256 AATGQWQTIIDGKKYFNTNTSIASGTGTIINGKHFFYFNTDGIWQIG---VFKGPNGEYFA 312
QY 1459 -----DVNSGDKKV-----NGFFTTGDNAMYADGQGNLAK---GRKSIDNQDLYFDPATGK 1507
Db 313 PANTDANNIEGQAIRYQNRFLYLDHNIYF-----GNNSKAVTGWQTINGNVYFMPDITAM 368
QY 1508 QVKGQLVSDGRNYYF 1523
Db 369 AAAGGLFEIDGVYIFF 384
RESULT 10
US-11-052-554A-217
; Sequence 217, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 217
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Streptococcus mutans UA159
US-11-052-554A-217
Query Match 3.5%; Score 290; DB 7; Length 726;
Best Local Similarity 22.4%; Pred. No. 4.1e-08;
Matches 142; Conservative 76; Mismatches 209; Indels 206; Gaps 28;
QY 946 EVVTAKRTNSYGNPTFDAYIN---NALYATVTKSGSYQAYGGAFLDELKAKYPDMF 1002
Db 38 ETVITEQROTSKINASSQVENQTSNQVEA-KTDSANKDPQEKTSVATDA-----PSMN 91
QY 1003 TVNMISTGKPIDPSIKIQWEAKYFNGTNVLKGAGYVLSDDATGKYFTVNEENGDFLPAS 1062
Db 92 SANNSQSD-----KQ-----NTVNEISSDSQQTKTDRQTD-----LPQN 126
QY 1063 FTGDQNAKTGFYDGTGMAYYSTGNK-----AVNSFIYEG-GHYYYFDKGHHMTGSKYA 1117
Db 127 SFKQQSAAHYKM-----TTEAEKTPSHSINTFVNDGNGWYVYLGADGRNVTGSH-- 174
QY 1118 EDGNDYFIFLPGIQMRDAIYQDAQNSYYIGRTGILYKGDNWPFPDNNANKTVFRYFD 1177
Db 175 ----- 174
QY 1178 ANNVMAIGYRNMYGOTYYFDENGFOAKGQLLTDDKGT--HYFEDNGAMAKNFPVNVGDDW 1236
Db 175 -----TIG-----GKTMVFAQDGKQVKGAFQADSGNKHYYDRDSEGMTNFRVNDQGNW 224
QY 1237 TYMDGNGNAVKQYYPVNNQILYFNPEYVQVKGQFIDTAQGRTSYYDANSAL KSSGFFT 1296
Db 225 YYLNDGVPVTSGLITVNGQSLYFNSD--GSQVKGNFV--BEDGSLRYDYDKNSGDLRLKTSRT 282
QY 1297 PNGSDWYEAENGYYVYKGFQVAENQDQWYFDTQTTGKQAKGAAKVDGRDLFYNPDSGVQV 1356
Db 283 INGVNYQFDNDG-----NARAIDKIE-----VV 305
QY 1357 KGFATDESNTSFYHGDNGDKVVGFFTTGNNAWYADNNGNLVKGFQEIIDGKWHFDE 1416

Db 1674 SIIFNGNDTIDPSKYQAGALIFASNGVSNINITLTNATNGLSNAGLNNSVQKGEICINL 1733
QY 1187 RNM-----YQOTYFDENGFOAKGQLTDDK-----GTHYFDE 1219
Db 1734 ANCPFTKNSSPANSVTPPTNESVHANNFTFLGTIISNGAIDLSQVNNNSVIGTILNE 1793
QY 1220 D-----NGAMAKNKFVNVDWYMDGNGNAVKQYFVNQIILYFNPETGVQVKGQF--- 1271
Db 1794 NATLOANNLTITAFNNASNSTANIDGN-----FTLNOQATLSTNASGLNVMGNFSY 1846
QY 1272 -----ITDAQO-----RTSYDANS- 1286
Db 1847 GDLVFNLSHVSIIINTQTATIMANNPLIOFNASSKEVGTYTLIDSAKAIYGVNNO 1906
QY 1287 -----GALKSGSFTPNNGSWYAEAGVYVYK-----GFKQVAENQDOWY 1325
Db 1907 ITGSSSLDNLKLYALIDINGKHMVMTDGLTYNGQAVSVKDGGLVVGFK---DSQNYI 1963
QY 1326 YFDQTTGKQAKGAKVD-GRDLYFNPDSG-----VQVKGDPATDESNGTSFYHGNGDK 1378
Db 1964 YTSILYNK-----VKIAVSNDPINNPOAPTLKQYIAIQGVQSVDSIDQAG---GNQAIN 2015
QY 1379 VVGGFPTTGN---AWYADNN-----GNLVKGFQFIDCKWYHFEV-----T 1418
Db 2016 WLNKIFETKGSPLFAPYLYESHSTKDLTTIAGDIANTLEVIANPNFNDATNLIQNTYT 2075
QY 1419 GQAKGA-----ALVNGQQLYFVDVSGIQVKGDFVTDG 1451
Db 2076 QQMSRLAKLSDTSTFARSDFLERLEALKNKRFPADAI PNAMDVILKYQSRRVKNVWATG 2135
QY 1452 QGNTSY-----YDVNSG-DKXVNGFPTTGNWYADGQGNLAKGRKSIDNODLY-- 1500
Db 2136 VGGASPISGGTGLYGINVGYDRFKIGVIVGGYAAVYSGFHANITQSGSSNNVNVGVYSR 2195
QY 1501 -----FDPATGKQVQQLVSDIGRNVYFDS-----G 1526
Db 2196 AFIKSELWMSLNETWYKNTFINSYDP-----LUSIINQSYRTDTWTTDAKINYG 2246
QY 1527 SGNMAKNREV----RIGDQWYFNGNDG 1549
Db 2247 YDFMFKDKSVIFKQVGLSYVYIGLSG 2273

RESULT 13

US-11-052-554A-228
; Sequence 228, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 228
; LENGTH: 338
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae R6
US-11-052-554A-228

Query Match 3.2%; Score 267; DB 7; Length 338;
Best Local Similarity 25.9%; Pred. No. 2.4e-07;
Matches 97; Conservative 53; Mismatches 126; Indels 98; Gaps 20;

QY 1145 YYYG-RTGILYKGNWYFFVDPNNANKTVFYFDPANNVMAIGYRNMYGQTYFDENGFQ 1202
Db 15 FFFGLLATNTVFANTTGGREVDKNNR-----KYYVKDDHKAIFYMHKIDKTYFFGDIGEM 69

QY 1203 AKQQLTDDKGTHTYFDE--DNGAMAKNKFVNVDWYMDGNGNAVKQYFVNQIIL--Y 1258
Db 70 VVCWQYLEIPGTGYRNLFDNPVNE--IGLQEKWYFQGDGALLE--QTDKQVLEAK 123
QY 1259 FNPETGVQVKGQFITAQGRTSYTDANSALKSSGFTPNNGSDWYIYAE-- 1307
Db 124 TSENTGKYVGEQYPLSAEKRTYTFD-NNYAVK-TGMIYEDG-NWYILNKLGNFGDDSYNP 180
QY 1308 ---GYVYKGFQK---VAENQDQ-----WYFDDQTTGKQAKGAKVDGRDLYENPDSGVQV 1356
Db 181 LPGEVAKWTDQFHTIIDIDRSKPAWYILD-ASGRMLTDWQKVGKWTYF----- 231
QY 1357 KGFDPATDESNGTSFYHGNDGDKVVGGFPTTGNNAWYADN-NGNLVKGFQFIDCKWYHFD 1415
Db 232 -----GSSGSMATGKYYVRGK--WYLDNKNKGDMKTGWQLGNKWKYLL- 272
QY 1416 EVTGOQAKGAALVNGQQLYFVDVSGIQVKGDFVTDGQNTSY--DVNSGDKKXVNGFPTTG 1474
Db 273 ---RSSGAMVTGWYQ-----DGLTWYILNAGNGDMKTGWFOVNG 308
QY 1475 DNAYVYADGQGNLA 1488
Db 309 ---KWYIAYSSGALA 320

RESULT 14

US-10-793-626-1780
; Sequence 1780, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1780
; LENGTH: 1155
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-1780

Query Match 3.2%; Score 267; DB 6; Length 1155;
Best Local Similarity 19.3%; Pred. No. 1.2e-06;
Matches 294; Conservative 193; Mismatches 496; Indels 540; Gaps 78;

QY 1 MEKLLHYKLHKYKHHV---TIIVASIGLVSLVSGAGTVAEDKVAANDTTAQTAVGVDTGQ 57
Db 9 INKRVDFLSNKNVYSIRKFTVGTASI---LVGA----- 39
QY 58 DOATTNDANTNTTDDTADQSANVQDQAGSDQSNQDQAKQDQATANTDRNQADNSQTDNN 117
Db 40 -----TLMFGAADNEAKAEDNQLESASKEEQKGRDNESSKLNQVD---LDNG 85
QY 118 QATDQATSPATDGTSVORRDAANVATAADQEQTAPSEQEKSAALSLDNVKLDIGKYYV 177
Db 86 SHSSEKTTNVNNAVEVKVVEAPTTSDVSKPKANEAVVNTSEKTP----- 129
QY 178 QADGSYKKNFAITVNGQMLYFSDTGALSSTSYSGTQTTNLVDDFSSHNKAYDSTAKS 237
Db 130 -----KITEAPTVEESI---AETPKSTTQ-----QDSTEKN 159
QY 238 FELVNGYLTANSWYRPAGILRNGQTWEASNENDLRPVLMSWVDPDKDTQVAYVNNMKYLS 297
Db 160 -----NPSL-----KD-----N 166

QY 298 ANETEVNTSQQD--LNKEAQSIQTKIEQKITSNSTQWLTAMEAFVAAQPKNMSTE 355
Db 167 LNSSSTTSKSTDEHSTKQAO--MSTNKSNLNDTNSPTQSEKTSQA----- 212
QY 356 NFNKGDLGGALLYNSDLTPWANSDYELLARTPTQDGTGKYFTEGGEGGYEFLLSND 415
Db 213 -----NDSTNQAPSQOLSKSEQKYTKFNDEP-----TQD 248
QY 416 VDSNPNVQAOQLNQLHYLMNMGDI VMGDKDANFQGRVDAVDNVMNADI--LQVYSNYPK 473
Db 249 VEHTTTKLKTPSIST-----DSSVNDKQ--DYTR-SAVASLGVDNTEATNAVR 296
QY 474 DNYKVTDSANALAHISILEAW--SLNDNQYNEDTNGTALSNDSSRLTSLAVLTKPOCQ 531
Db 297 DNLDLKAAREQINERAIIEALKKDFSNPDYGD--TFLALNTSOSKNS--PHKSASP 350
QY 532 RIDLSNLISEVNKERANDTAYGDTIPTYSFVRAHDSVQTVIAKIVKEKI----- 582
Db 351 RNNLMSLAAPNSGKVNNDKV-KITNPTLSLNS-KNNHANNVWPTSNEQFNLKANYELD 408
QY 583 DTNSDGYTFTL-----DQLKD-----AFKIYNEDMAKVNKTYTHY-- 617
Db 409 DSIKEGDTFTIKYQVIRPGGLEPAIKQLRSKDGSI VANGVYDKTTNTTTPNYVD 468
QY 618 ---NTPAAVALL-----LSMESVP-----RVYGGDLYTDDQYMAKSPYYDAI 659
Db 469 QYCNITGSPDLIATPKRETAIKDNQNPMEVTIANEVWKDFIVDYG--KK-----DNT 521
QY 660 ATMLQGRIAVVGSGGSEEVHKGNNQ-----ILSSVRYGO-----DLMSAD 701
Db 522 TTAANVANVDNNKHNVEVY-LNQNQNPKYAFSVKNGKFIPEGEVKVYEVTDNAMV 580
QY 702 DTQGTDLRSRTGLVTLVSNPDNLDLGGDSLTVNMGRAHAN-----QAYRPLILG----- 750
Db 581 DSNFPLNSSNKDVTSQPTPKVSADGTRVDINFARSMANGKXYIVTQAVRPTGTGNVYT 640
QY 751 ----TKDG---VOSYLKDSSTNIVKTYTDANGNLTF TADDIKGYSVTVMGSGYLAVVVP-- 801
Db 641 EYWLTRDGTNTNDFYRGTKSTTVTY--LNGSSTAQDN-----PTYSLGDY--VMLDKKN 692
QY 802 -GAKGQDVRVAADTNQKADGKSLKTSAAQDSQVIYEGFSNQ-----DF----- 845
Db 693 NGVQDDDEKLGAVVYTLKDSNNRELQRTVTDQSGHYQFNDLQNGTYTFVEFALPDNYTSP 752
QY 846 -ANNADYTNKKIAENADFFKLGITSFEMAPQYVSATGDSFLDSIIQNGYAFSDRY---- 901
Db 753 PANN--INDAIDSQGE--RDGTRKVVVAKGTINNADNMVTD-----GFYLTPKYNVG 802
QY 902 DLAMSNNKYGSKDDLANALKALHANGIOAIADWVPDQIYQLPGEVVTAKRTN--SYGN 959
Db 803 DYVWEDTNKDGIQDD-----NEKGISNVK-----VTLKKNKGDTIGT 839
QY 960 PTFDAYINNALVATNTKSSGSDYQAOYGGAFGLDELKAKYPMFTVMNISTGKPIDPSTKI 1019
Db 840 TTLD-4-SNGKYEFTGLENG-DTIEF-----ETPEGVTPTKQNSGSD----- 878
QY 1020 KOWEAKYFNGTNVILGKAGVYVLSDDATGKYFTVNEGDFLPASFTGDQNAKTGFYDGTG 1079
Db 879 ---EGKDSNGTKTT-----VTVKADNK--TID-----SGFYKP--- 907
QY 1080 MAYYSTSGKANVNSFYEGGHYYFD--KQGHMVTSYKAEDGNDYVFLPGLQMRDAIY 1137
Db 908 -----IYNLGDYVWEDTNKDG-----IQDDSE-----KGISGVKVTL 939
QY 1138 QDAQGSYVYGRGILYKGNWYPFVDPNNAKTVFPRYFDANNVMAIGV---RNMYGOTY 1194
Db 940 KQKNQNA--IGITTTDASH--YQFKGLENGSTV--EFTPS-----GTPPTKANSQQDI 989
QY 1195 YFEDENGFOAKGQL-----LTDDKRGTHYFDEDNAGAMAKNFVNNGDDWYMDGNGNAVKQ 1249
Db 990 TVDSNGITTTGIINGADNLITIDSGFY-----KTPKYSVG-DYVWEDTNKDG IQDD 1038
QY 1250 YPVNNQILFNPETGVQV-----KGQFI-----TDAQRTSYYYDANSALKSSGFFTPNGS 1300

Db 1039 ---NEKGI-----SGKVTLKDEKGNILSTTTTIDENGKQFNDLSGNY-IHFEPKPEGM 1089
QY 1301 DWYAEANGYVYKGFQKQVAENQDQWYFYDQTTTGKQAKGAQKVDGRDL----- 1346
Db 1090 TQTTANSQ-----NDDE-----KDA-----DGEDVRVTITDHDDFSIDN 1123
QY 1347 -YFNPDSGVQVKGDFATDESNT 1368
Db 1124 GYFDDSDSDSDADSDSDSDS 1146
RESULT 15
US-11-052-554A-372
; Sequence 372, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: 2005-02-07
; PRIOR FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 372
; LENGTH: 1767
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae R6
US-11-052-554A-372
Query Match 3.2%; Score 265; DB 7; Length 1767;
Best Local Similarity 19.4%; Pred. No. 2.7e-06;
Matches 287; Conservative 171; Mismatches 519; Indels 504; Gaps 77;
QY 371 TNSDLTP-WANSDYRLNARTPTQDGTGKYFTEGGEGGYEF-LLSNDVDNNSPNVQAEQL 428
Db 117 TPKTVTPEWQTEKK-----EQQT---VTIREKGVRYNQLSSAQNDN----- 158
QY 429 NOLHYLMNMGDILVMGDKDANFD--GURVDANVDNADLLQVYSNYFKDN----- 475
Db 159 -----AGKPALPEKGLTVDANGNAIVDLT-----FKDDSEKGSRFGVF 198
QY 476 YKVTDSANALAHISILEAWSLNDQYNEBT-----NGTALSIDNSSRLTSLAVLTKPQG 530
Db 199 LKPKDTKNVFGYD-KQGFWM---EYKSPTTSTWVRGSRVAAPETGNTNRLSITLKS DG 254
QY 531 QRIDLSNLISESVNKERANDTAYGDTIPTYSFVRAHDSVQTVIAK-----IVKEK 581
Db 255 Q-LNASN-----NDVNLFDITVLPAAVNDHLKNEKILLKAGSYDDERTVVSVK 302
QY 582 IOTNSDGYTF-----TLQDLKDAFKIYNEDMAKVNKTYTHYINPAA--YAL----- 625
Db 303 TD-NOEGVKTEDTPAKETGPEVDBSKVTYDTIQSKVLKAVIDQAPPRVKESYLNIGHTLP 361
QY 626 -----LLGNMESVPRVYVYDLYTDDGQYMAKSPYYDAIATMLQGRIAVVGSGQSE 676
Db 362 GOVQFNOVFINHRIITPEVTYKKNETTAETAEYLMKLURDDAHLINABMTVRLQVVDNQLHF 421
QY 677 EVHKVNGNNQI-----LSSVRY-GODLMSADDTQ-----GTDLSRTSGLVTL 717
Db 422 DVTKIVNHNQVTPGQKIDDERKLLSSISFLGNALVSVDQTCAGKPDGATMSNN-----TH 477
QY 718 VSDNPNLDLGGDSLTVNMGRAHANQAYRPLILGTGDKGQV-SYLKQSD--TNIVKYTDANG 774
Db 478 VSGDDHIDVTNPMKOLAKGYMFGFVSTDKLAAGVMSNSQNSYGGGSDNMTRLTAYKETVG 537
QY 775 NLTFATDADIKGYSTVMDSGYLAVVVPVGAQKQDVRVAADTNQKADGKSLKTSAAQDSOV 834

Search completed: February 11, 2006, 20:58:39
Job time : 24.8432 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.
OM protein - protein search, using sw model
Run on: February 11, 2006, 19:15:50 ; Search time 32.0752 Seconds
(without alignments)
4661.567 Million cell updates/sec
Title: US-10-797-821-38
Perfect score: 8237
Sequence: 1 MEKKLHYKLHKVKKHWTTIA.....FVRIGDQWYFGNDGAATNL 1554
Scoring table: BLO62UM62
GapOp 10.0 , Gapext 0.5
Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : PIR 80.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query					Description
No.	Score	Match	Length	ID		
1	4014	48.7	1599	2	S22737	glucosyltransferase
2	3818	46.4	1518	2	A44811	glucosyltransferase
3	3635	44.1	1431	2	A45866	dextranucrase (EC 3.2.1.11)
4	3472.5	42.2	1592	2	A38175	glucosyltransferase
5	3434.5	41.7	1475	2	B33135	gtfB protein precursor
6	3429.5	41.6	1577	2	T30858	glucosyltransferase
7	3264	39.6	1449	2	T30857	glucosyltransferase
8	3243	39.4	1449	2	T30552	glucosyltransferase
9	3161.5	38.4	1375	2	J70345	dextranucrase (EC 3.2.1.11)
10	3080.5	37.4	1365	2	A14483	glucosyltransferase
11	2964.5	36.0	1508	2	T31098	probable dextranucrase
12	2799	34.0	1290	2	JC5473	dextranucrase (EC 3.2.1.11)
13	654.5	7.9	2817	2	B97033	uncharacterized protein
14	545.5	6.6	2710	2	A37052	toxin A - Clostridium
15	498.5	6.1	2364	2	I40884	cytotoxin L - Clostridium
16	464.5	5.6	2178	2	S5805	alpha-toxin - Clostridium
17	463	5.6	2366	2	S10317	toxin B - Clostridium
18	454	5.5	648	2	S10869	enterotoxin A - Clostridium
19	443.5	5.4	2367	2	S70172	toxin B - Clostridium
20	430.5	5.2	563	2	A37184	glucan-binding protein
21	357.5	4.3	2334	2	S32920	cell wall-associated
22	331.5	4.0	329	2	A52221	dextranase inhibitor
23	331.5	4.0	1829	2	E81086	iron-regulated protein
24	324.5	3.9	1829	2	S35027	cytotoxin RTX homolog
25	323	3.9	2167	2	AF1489	cell wall-associated
26	314.5	3.8	1463	2	T30290	AAS surface protein
27	298	3.6	1902	1	B44858	lactococpin (EC 3.4.1.1)
28	285	3.5	1385	2	D89824	hypothetical protein
29	281.5	3.4	340	2	G95043	choline binding protein

30	280.5	3.4	3705	2	AD0123	probable autotrans
31	278	3.4	1315	2	T28679	fibrinogen-binding
32	277.5	3.4	2044	2	AB1180	probable peptidogl
33	276.5	3.4	721	2	C97980	endo-beta-N-acetyl
34	276.5	3.4	2399	2	H71879	toxin-like outer m
35	276	3.4	1305	2	H41662	150K mating aggreg
36	275	3.3	2529	2	B64515	toxin-like outer m
37	275	3.3	4936	2	AH2515	hypothetical prote
38	273.5	3.3	1904	2	T13256	tail-host specific
39	273	3.3	3844	2	T18402	asparagine/asparta
40	272	3.3	1441	2	B86807	hypothetical prote
41	270	3.3	1335	2	T30211	autolysin B - Stap
42	269.5	3.3	1821	2	AG2335	hypothetical prote
43	268.5	3.3	2348	2	AD1841	hypothetical prote
44	268.5	3.3	13055	2	TL6580	hypothetical prote
45	267	3.2	338	2	A97914	choline-binding pr

RESULT 1
S22737
glucosyltransferase (EC 2.4.1.-) S - Streptococcus salivarius
C:Species: Streptococcus salivarius
C:Date: 16-Sep-1992 #sequence revision 16-Sep-1992 #text_change 09-Jul-2004
C:Accession: S22737; S28810; B44811; S22727
R:Jacques, N.
submitted to the EMBL Data Library, March 1992
A:Reference number: S22726
A:Accession: S22737
A:Molecule type: DNA
A:Residues: 1-1599 <JAC>
A:Cross-references: UNIPROT:Q00599; UNIPARC:UPI00000BEF34; EMBL:Z11872; NID:g47530; PID: P1D
A:Experimental source: ATCC 25975
R:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.
J. Gen. Microbiol. 137, 2577-2593, 1991
A:Title: Molecular characterization of a cluster of at least two glucosyltransferase ge.
A:Reference number: A44811; MUID:92148377; PMID:1838391
A:Accession: S28810
A:Molecule type: DNA
A:Residues: 1-51 <Gif>
A:Cross-references: UNIPARC:UPI000017027C; EMBL:Z11873
C:Genetics:
A:Gene: gtfK
C:Keywords: glucosyltransferase; hexosyltransferase
F:1456-1475/Domain: cpl repeat homology <CPR>

Query Match 48.7%; Score 4014; DB 2; Length 1599;
Best Local Similarity 51.7%; Pred. No. 1.5e-172;
Matches 851; Conservative 215; Mismatches 430; Indels 150; Gaps 40;

Qy 1 MEKKLHYKLHKVKKHWTTIAVASGLVAGTUSA-----EDKVANDTTAQATVGDV 54
Db 1 MENKRYKYLHKVKKQWVTLAVASVALATIVGGSVATSSLASAETNNNGSPSTTVG 60
Qy 55 TG-----QDOATTNTDANT-NTTDTD---TADQSANTN-QDOAGSDQSNQDOAQD 103
Db 61 TNPVVEKVGTTTEVANTSNATTTTAEAVTADKPAGTTVPNSGT-----T 106
Qy 104 TDRNQA--DNSQTDNNQATDOATSPATDGTSVQRDAAANVATAA--DQEGQTAPSEK 160
Db 107 SDRAAAVEAVEKPETTAKPEVATKPTATTS--EVAANAGVAAPTTEKSELSAEIKA 163
Qy 161 ALSLDNVKL-IDGKYYYVQADGSKYKKNPAITVNGQMLYFSDGTGALSSTSTVSPQGT 219
Db 164 AVSLDNIKKEDGKGYLLLEDGSHKKNPAITVNGQVLYFD-ENGALSSTSTVSPQGT 222
Qy 220 LVDDFSSHNKAYDSTAKSFELVNGVLTANSVVRPAGILRNGOTWEASNENDLRPVLS 279
Db 223 LVTDFTKNNAAVYDSTAKSFELVDGVLTAADSVYRPRKLEAGTTWKASTEKDFRPL 282
Qy 280 PDKDTQVAVVNMNKKYLS-ANETE--VTNETSQVDLNKEAQSIQTKIEQKITSDNST 336

Db 417 DPHMGDFELLANDVDLSNPVQAEQNLQIHLYLNMWGSIVMGDKDANFGIRVADVNV 476
Qy 461 NADLQVSYNYFKDNYKVTDSANALAHISILEAWSLNDNQYNEDTNGTALSIDNSRLT 520
Db 477 DADMLQLYTNVFRYYGVNKSANALAHISVLEAWSLNDNHNDKTDGALAMENKORLA 536
Qy 521 SLAVLTK-----OPQORDLSNLSSESVNKERAND-----T 551
Db 537 LFLSLAKPIKERTPAVSPLYNNTFTTQRDEKTDWINKDGSKAYNEDGTVKQSTIGKYNE 596
Qy 552 AVGDTPITYFVRAHDSVOTVIAKIVKEKIDTNSDGYTFTLQDKAPKIYNEHMAKVN 611
Db 597 KYGDASGNVYFIRAHNDNVODIAEIIKKEINPKSDGFTTIDAMEKQAFIYNKMDLSSD 656
Qy 612 KTYTHYNIPAAAYALLSNMBSVPRVYDLYTDDGQYMAKSPYDAIATMLQCRATYVS 671
Db 657 KKYTLNIPAAAYAVMLQNMETITRVYDLYTDDGHYMETKSPYDITVNLMSRIKYS 716
Qy 672 GQOSEVH-----KVGNNQILSSVRYGQDLSMSADDTQGTDLRSRTSLVTLV 718
Db 717 GGAQORSYMLPTDGRKMDNSDVELYRTNEVYTSVRYGKDIMTANDTEGSKYSRTSGQVTLV 776
Qy 719 SNDPNLDLGGDS--LTVNMGRAHANOAVRPLILGTGQVQSYLKDSD---TNIVKYTDANG 774
Db 777 ANPKLNDQSAKLANVEMGIHANQKTRALIVTAGIKNFTSDADAIAGYKYKETSNG 836
Qy 775 NLFTTADDIKGYSTVDMSGYLAVVVPVGAKDQDVRVAADTNQADGK-SLKTSAAALDSQ 833
Db 837 VLTFGANDIKGYETFDMSGFVAWVVPVASDNDQDIRVAPSTEAKKEGELTLKATEAYDSQ 896
Qy 834 VIYEGFSNFQDF--ANNADYTNKKIAENADFPKLGITGSFEMAPQVVSATDGSPLDSII 891
Db 897 LIYEGFSNFQTPDGSDPSVYTNKIAENVDLFKSWGVTSEFMAPQFVSADDTGLDSVI 956
Qy 892 QNGYAFSDRYDLAMSNNKYGSKDLANALKALHANGIOALADWVPDQIYOLPGBEVTA 951
Db 957 QNGYAFADRYDLAMSNNKYGSKEDLRLDALKALHAGIOALADWVPDQIYOLPGBEVTA 1016
Qy 952 KRTNSYGNPTFDAYINNALYATNKS GSDYQAYGCAFDELKAKYPMFTVMISTGK 1011
Db 1017 TRTDGAGRIADAIIDHSLYVANSKSGKDYQAKYGEFELAEIKAKYPEFKVMISTGK 1076
Qy 1012 PIDPSTKIKQWEAKYFNGTNVLKGAGYVLSDDATGKYFTVNBENGDFLPASFTGDQNAKT 1071
Db 1077 PIDDSVKLKQWKAERYFNGTNVLERGVGYYVLSDEATGKYFTVTKEGNFIPLQLTGKEKVT 1136
Qy 1072 GFYDGTGMAYYSTGNKAVNSFIYEGCHYFFDKDGHMVTGSKYKABDNDYF-FLPNCI 1130
Db 1137 GFSSDGKGYFTGTSGTQAKSAFVTFNGNTYFPDARGHMTNSEYSPNGKDVYRFLPNCI 1196
Qy 1131 QMRDALYQDAQGSYVYGRGTILYKGDNWPFFVDPNNANK-----TVFRYFDANNVMA 1183
Db 1197 MLSNIFYIDANGNTLYNKSQOMYKGG--YTFEDVSETDQKESKVKVFRFTNEGVA 1254
Qy 1184 IGRNNYMGTYTFDENGFOAKQLLTDKGTTHYFDESDGAMAKNFVNNYGDVMDGNG 1243
Db 1255 KGVTVTDGTFQYFGEDGFOAKDLVTFKGKTYTFDAHTGNGIKDTRNNGKRWYFDANG 1314
Qy 1244 NAVKGQYFVNNQILFNPETGVQVKGQFTIDAGRTSYTDANSALKSSGFTTPNGSDWY 1303
Db 1315 VAATGAQVINGOKLYFN-EDGSQVKGGVVKNADGTSYKKEGFGELVTFNEFTTIDGNVWY 1373
Qy 1304 YAENGVYVYKFPQVAENQDQWYFDTQTCQAKGAUKVDGRDLYFNPDGSGOVKGDPAFD 1363
Db 1374 YA-----GANGKTVTGAQVINGQHLVFNAD-GSOVKGSGVKN 1409
Qy 1364 ESGNTSFYHGDNDKVVGGFTTGNNAWYADNNGNLVKGFOEIDGKWHFDEVTCQAK 1423
Db 1410 ADGTVSKYNASTGERLTNEFTTIDGNWYIIGANGKSVTGEVKI----- 1453
Qy 1424 GAALVNGQOLYFDVDSGQVKGDFVTDGQNTSYDVNSGDKKXNGFFTTGDNAMYADG 1483
Db 1454 -----GDDTYFPAKDGQVKGQTVSAGNRSIYYGSGKRAVSTWIEIQPVVYVYFDK 1507

Qy 1484 QG 1485
Db 1508 NG 1509

RESULT 3

A45866
dextranucrase [EC 2.4.1.5] precursor - Streptococcus mutans
C:Species: Streptococcus mutans
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Mar-2004
C:Accession: A45866

R;Honda, O.; Kato, C.; Kuramitsu, H.K.

J. Gen. Microbiol. 136, 2099-2105, 1990

A:Title: Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the glucosyl

A:Reference number: A45866; MUID:91100958; PMID:2148600

A:Accession: A45866

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1431 <HON>

A:Cross-references: UNIPARC:UPI000017AC5C; GB:M29296

C:Keywords: glycosyltransferase; hexosyltransferase

F;1127-1146/Domain: cpl repeat homology <CP1>

F;1192-1211/Domain: cpl repeat homology <CP2>

F;1257-1276/Domain: cpl repeat homology <CP3>

F;1277-1297/Domain: cpl repeat homology <CP4>

F;1321-1340/Domain: cpl repeat homology <CP5>

F;1341-1361/Domain: cpl repeat homology <CP6>

F;1385-1404/Domain: cpl repeat homology <CP7>

Query Match

Best Local Similarity 44.1%; Score 3635; DB 2; Length 1431;

Matches 749; Conservative 218; Mismatches 422; Indels 112; Gaps 32;

Qy 1 MEKKLHYKLUHKKVHWVTIAVASIGLVSVAGTVAEDKAVANDTTAQAATVGYDTGQ-- 57
Db 1 METKRYKMKHKVHWVTIAVAS-GLITL---GTTLGSVSAETSEQTSKVTQKSED 56
Qy 58 DOATINDANTNTDITDQOSANTWQOAGSQSNNOQAKQDT-----ANTDR 106
Db 57 DKAAESSQTDAPKT----KQAQTEQTOAQS-QANVADTSTITKSTPSQNIITQANSDD 111
Qy 107 NQADNSQTDNNQATQATSPATDGTSVQRDDAANVATA-ADQEGQATPSEQEKSAAL-SL 164
Db 112 KTVYTKSEAGTSEERTKQSEBAQTASSQALTQAKBELTKQRTAAQENKNPVDLAAI 171
Qy 165 DNVLKIDGKYVYQADSGYKKNFAITVNGQMLYFSDTGALSTSTYSFSQGTTLNLDVF 224
Db 172 PNVKQIDGKYVYIGSDGQPKKFNALTNNKVLVFDKNTGALTDTSQYQFKQGLTKLNDY 231
Qy 225 SSHKAYDSTAKSPFLVNGYLTANSWYRPAGLRNGQTWAEANENDLRPVLMSWPDKOT 284
Db 232 TPNQIVNFENTSLKETIDNYVTADSWYRPKDILKNGKVTWASSSEDLRPLMSWPDKOT 291
Qy 285 QVAYVYNNKYLANSATEVTNETSOVDLNKEAQSTQTKLEQKITSNSTMWLTAMEAFV 344
Db 292 QIAYLNYNNQOGLGTGENVTADSSQESLNLAAGTQVQKTIETKISQTOQWLRDIINSFV 351
Qy 345 AAQPKNNMSTE---NFKNGDHLQGGALLYTNSDLTPWANSYRLLNRTPTQDDGTKKYFT 401
Db 352 KQPNWNSQTESDTSAGEKDHLOGGALLYSNSDKTAYANSYRLLNRTPTSGTKPKYFE 411
Qy 402 EGGEGGYFLLSNDVNDNPNVVOAEQLNQLHYLMWGDIVMGDKDANFGVIRVADVNVN 461
Db 412 DNSSGGYDFLLANDIDNSNPVVOAEQLNQLHYLMWGSIVANDPEANFGVIRVADVNVN 471
Qy 462 ADLLQVSNYPKONKVKVTDSEANALAHISILEAWSLNDNQYNEDTNGTALSIDNSRLT 521
Db 472 ADLLQIASDYLKAHYGVDKSEKKNINHLISLEAWSNDPQYNKOTGAQLPINDKLRLSL 531
Qy 522 LAVLTQPCQRIDLSN-----LISESVNKERANDTAYGDTIPTYSFVRAHDSVQ 571

Db	532	LYALTR-PLEK-DASNKNEIRSGLEPVITNSLN-NRSEAGKNSERMANYIFIRAHDSVEQ	588
Qy	572	TVIAKTVKSKIDTNSDGYFTTLDQKDAFKIYNEDMAKVNKYTHYNIIPAAVALLSNME	631
Db	589	TVIAKIIKAQINPKTDGLTFTLDQKDAFKIYNEDMRQAKKYTQSNIPTAYALMSNKD	648
Qy	632	SVPRVYVGYDLYDDGOYMAKKSPYDAIATMLQGRITAYVSGGSEBHVKNVGNQ----	686
Db	649	STRLLYGYDMSYDDGOYMAKSPYDAITDLKARIKYAAGQDMKITVEGDKSHMDW	708
Qy	687	----ILSSVRYGODLMSADDTQGTDLRSRTSLGVLTVNSNDPNLDLG-GDSITVNMGRAHANQ	742
Db	709	YTGVLTSVRYGTGANEATD-QGSEATKTQGMVITSSNPSLKLNQNDKVIIVNMGAAHKQ	767
Qy	743	AYRPLILGTGQGVQSILKSDT-NIVKYTDANGNLFTPTADDTKGYSTVDMSGYLAWVPV	801
Db	768	EYRPLLLTTTKDGLTYSYDAAAKSLYRKTNDKGLVFDASDIQGYLNPQVSGYLAWVPV	827
Qy	802	GAKGODVRVAADTNQKADGKSLKTSAAALDSOVIVGPGSNFODPANNDADYTNKKIAENA	861
Db	828	GASDNODVRAASNKANATQGVYESSALDSQLIYEGSFNQDFVTVKSDYTNKKIAQNV	887
Qy	862	DPFKLIGITSFEMAPQVYSATDGSFLDSIIQNGYAFSDRYDLAMSKNNKYGSKDILANAL	921
Db	888	QLFKSWGVTSEFAPQVYSESDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGQQDMINAV	947
Qy	922	KALHANGIOAIADWPQYIQLPGEVVTAKTNSYGNPTFPDAYINNALIYATNWKSSGD	981
Db	948	KALHKSIGIOVIADWPQYINLPGEVVTATRVNDYGEYRKDSKSEIKNTLYAANTKSNKGD	1007
Qy	982	YQAOYGAELDELKAKYPMFTVMNISTGKPIDPSTKIKQWEAKYFNGTNVLGKGAYVL	1041
Db	1008	YQAKYGGAFSELAAYKYPISFNRTQISNGKKIDPSEKITAWKAKYFNGTNILGRGVGYL	1067
Qy	1042	SDDATKYFTVNEGDFLPASFTGQONAKTGFYDGTGMAYYSTSGNKAVNSFIYEG-GH	1100
Db	1068	KONASDKYFELKGNOTYLPKQMT-NKEASTGFVNDGNGMTFYTSGYQAKNSFVQDAKN	1126
Qy	1101	YYFDFKDGHWVTGSYKAEQNDYFPLPNGIOMRDAIYQDAQNSYYTGTGILYKGNWY	1160
Db	1127	WYFFDNNGHMVYGLQOL-NGEVQYFLSNGVQLRESPLENADGSKNYFGHLGNRY-SNGY	1184
Qy	1161	PFVDPNNAKTVPFYEDANNMAIGYRNMVGTYYFEDENGFOAKGOLLTDDKG-THYFDE	1219
Db	1185	SP-----DNDSKWRYPDASGVMAVGLKTNGNTQYFDQDGYQVKGAWITFGSGDKRYFDD	1239
Qy	1220	DNGAMAKNFVN-VGDDWYMDGNGNAVKGQYVANNQILLYFNPETGVQVKGQFITDAQGR	1278
Db	1240	GSNGMAVNFANDKNGDWYILNSDGLALGVQTINGKTYFG-QDGKQIKGKIITD-NGK	1297
Qy	1279	TSYYDANSALKSSGFFTPNGSDWYYAENGYYVYKFKQVAENQDQWYFPQTGQAKGA	1338
Db	1298	LKYFLANSGLARNIPAT-----DSQNNWYFG-SDGVAVTGS	1334
Qy	1339	AKVDGRLYFNPDSGVQVKGDFATDESQNTSFVHGDNQDKVYGGFTTGNNAWYADNNG	1398
Db	1335	QTIAQKLYFASD-GKQVKGSFYT-YNGKVHYHADSGELQVNRFEADQGNWYLDNSG	1392
Qy	1399	NLVKGFQETDGKWHYHDEVTGQQAAGALVNGQQLYFVDVDSGIVQKGFVFTQDQGNYSY	1458
Db	1393	EALTGSQRI-----NQORVFF-TREGQVKQGVAYBERGLLRY	1430
Qy	1459	D	1459
Db	1431	D	1431

RESULT 4
A38175
glucosyltransferase precursor - Streptococcus sobrinus
C:Species: Streptococcus sobrinus
C:Date: 28-Aug-1992 #sequence_revision 28-Aug-1992 #text_change 15-Mar-2004
C:Accession: A38175

R;Abo, H.; Msumura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.
J. Bacteriol. 173, 989-996, 1991
A:Title: Peptide sequences for sucrose splitting and glucan binding within Streptococcus
A:Reference number: A38175; MUID:91123227; PMID:1704006
A:Accession: A38175
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1592 <ABO>
A:Cross-references: UNIPARC:UPI000012BCB2; GB:D90213; NID:g217032; PIDN:BAAL4241.1; PID
F:1093-1112/Domain: cpl repeat homology <CP1>
F:1227-1241/Domain: cpl repeat homology <CP2>
F:1287-1306/Domain: cpl repeat homology <CP3>
F:1330-1351/Domain: cpl repeat homology <CP4>
F:1352-1371/Domain: cpl repeat homology <CP5>
F:1402-1420/Domain: cpl repeat homology <CP6>
F:1465-1484/Domain: cpl repeat homology <CP7>
F:1513-1532/Domain: cpl repeat homology <CP8>

Query Match	42.2%;	Score	3472.5;	DB 2;	Length	1592;			
Best Local Similarity	45.1%;	Pred. No.	3.3e-148;						
Matches	741;	Conservative	254;	Mismatches	496;	Indels	153;	Gaps	39;
QY	1	MEKCLHYKLHKVKKHWVTIAVASIG-LVSLVGAGTVSAEDKQVANDTTAQATYGVDTGDOQ	59						
DB	1	MEKNVRFKMHKVKKHWVTLVASATMLASALGASVASADTDTASDDSNQTVV---TG-DQ	56						
QY	60	ATTNDA--NTNNTDTDTADQANTNQDQSDQSNNOQAOKDQTANTDRNQADNSQTDNN	117						
DB	57	TTNQATDQTSIAATATSEQASSTD--AATDQASAAEQTCQTASTD--TAAQTITNAN	111						
QY	118	QATDQATSPATDG-TSVORRDAANVATAADQEGQTAPSEQEKSAALSLDNVKLIDGKYY	176						
DB	112	EAKWPTSENQGFDEMLAEAKNVATA---ESDSIPSDLAK-----MSNVKQVDGKYY	163						
QY	177	VOADGSYKKNFAITVNGQMLYFSDTGALSSITSYSFSGQITNLVDD---FSSHNKAYDS	233						
DB	164	YDQDGNVKNFAVSGDKIYFD-ETGAYKDTISKVDADKSSSAVSQNAITIFAANNRAYST	222						
QY	234	TAKSPELVNGYLTANSWYRYPAGILRNGQTVKASNENDLRPLVMSWPKDQTVAVYVYN	293						
DB	223	SAKNFEADVNLITADSWYRPKSILKDGKTWTESGKDDFRPLMAWMPDTEKRTNVYMN	282						
QY	294	KYLSANEVEVNETSOVDLNKEAQSIQTKIEQKITSDNSTOWLRTAMEAFVAAQPKWMS	353						
DB	283	KVVGIDKT-YTAETSQADLTAAAEILVQARIEQKITSENNKWLREAISAFVTKQPMNGE	341						
QY	354	TENFNKGDLHQQGALLYTN-SDLTPWANSDYRLNRTPTQODGK---KYFTEGEGGY	408						
DB	342	SEK-PYDDHLQGALLFDNQDTLTPTQSNYRLNRTPTNQTSGLDSRFTYNPNPLGGY	400						
QY	409	EPFLSNVDNSNPVQAEOLNQLHYLMWGDIVMGDKQANFGVRVDADVNVNADLLQVY	468						
DB	401	DFLLANDVNSNPVQAEQLNWLHYLLNFGSIYANDADANFDSIRVDAEONVDADQIQIS	460						
QY	469	SNYFKDNYKVTDSSEANALAHISILEAWSINDQVNETNGTALSIDNSSRLTSLAVLTQK	528						
DB	461	SDYLKAAAYGIDKKNKNANNHVSIVEAWSNDPTPLYHDDGDNLMNMDNFKRLSLMLAKP	520						
QY	529	PQQRIDLNLNLSVESYKNERANDTAYGDTIPTYSFVRAHDSVEQTVIAKIVKEKIDTNSDG	588						
DB	521	TDVRSGLNPLIHNSL-VDRREVDDREVETVPISYSPARAHDSVEQDIIRDIKAEINPNSFG	579						
QY	589	YTFTLDLQKAPKIYNEDMAKVNKTYTHYNIIPAAVALLSNMESVPRVYVGYDLYDDGOY	648						
DB	580	YSFTQEEIDQAFKIYNEDLKKSKKYTHYNVPLSYTLTLTNKGSIPRVYVGYDMFTDDGOY	639						
QY	649	MAKSPYYDAITATMLQGRITAYVSGGSEBHVHKNVGNQILSSRVYRGODLMSADDTQGTDL	708						
DB	640	MANKTVNDALTESLTKARKMYVAGQAMQNYQI-NGEILTTSRVYRGKGLKQSD-KGDAT	697						
QY	709	SRTSGLVTLVSNNDPNLDLGGDSLTVNMGRAHANQAYRPLILGTGQGVQSILKSDTN---	765						
DB	698	TFTSGVGVMGQNPFLSDGKVVVALNMGAAHANQAYRALMVSTKDGVAITYATDADASKAG	757						

Db	688	GKALWATDT-GDRTTRTSVAVIEGNNPRLKASDRVVVNMGAHKQAYRPLLLITD	746
QY	753	DGVQSYLKSD--TNIVKYTDANGNLFTTADDIKGISTVDMSGYLAWVPVGA--DGQDV	809
Db	747	NGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKGVANPOVSGYLVGVVPGAALIKMFAL	806
QY	810	RVAADTNOKADGKSLKTSAAALDSOVIYEGFSNFDQFANNADYTNKKIAENADFKKLGI	869
Db	807	RLARPHQQA---SVHQNAALDSRVNFEFSNFAFATKEEYTNVIAKNVDKFAEWGV	863
QY	870	TSPFMAPOVVSATDGSFLDSIIQNGYAFSDRDYDLAMSKNKKYGSKDILANALKALHANGI	929
Db	864	TFEWAPOVVSATDGSFLDSVQNGYAFTRDYDLGSLSKPNKYGATADDLVKAIKALHSKI	923
QY	930	QAIADWPDOIYQLPGEVVTAKRTNSYGNPTFDAYINNALYATNTKSGSYQAYGGA	989
Db	924	KVMADWPQWYAFPEKEVVTATRVVDKYGTPVAGSQIKNTLVVDGSKSGDKQOAKYGA	983
QY	990	EJDELKAKYPDMFTVNMISTGKPIDPSTKI KWEAKYFNGTNVLGKAGYVLSDDATGY	1049
Db	984	FLEELQAKYPELFARKQISTGVPMDSVKIKOMSAKYFNGTNILGRGAGYVLKQOATNTY	1043
QY	1050	FTVNEGD--FLPASFTGQNAKTFYDGTGMAYYSTSCNAVNSFIYEGGHYYFDDK	1107
Db	1044	FNISDNKEINFLPKTLL--NQDSQVGSYDGKGVYYSTSGYQAKNTFISEGDKWYFDDN	1102
QY	1108	GHMVTGSYKAEDGNYYFLPNGIQMRDAIYQDAQNSYYGRTGILYKGDNNYPPVDPNN	1167
Db	1103	GYMVTGA--QSINGVNYFLSNGLQLRDAILKNEEDGTYYAGNDGRYE--NGYQFM----	1156
QY	1168	ANKTPRYFDANNVMAIGYRNMYGTYFYEDENGFOAKGQLLTTDDKG--THYFDEDNGAMAK	1226
Db	1157	--SGVWRHFN--NGEMSVGLTVIDGQVQFYDEMGYQAKGFVTTADGKIRYFPKQSGNMYR	1213
QY	1227	NKPV--NVGDWYMDGNGNAVKGOYPVNNOLLYFNPEGTGVQVKGFITDAQGRTSYDAN	1285
Db	1214	NRFIENEENKWLVLGEDGAATVGSQTINGQLYFR--ANGVQVKGEFVTDHGRISYDGN	1272
QY	1286	SGALKSGSFPFTNGSDWYAEYGVYKPKQVAENQDOWYFYDQTTGQAKGAAKVDGRD	1345
Db	1273	SGDQIRNR-----VRNAQQGFYFD--NNGYAVTGARTINGQL	1309
QY	1346	LYENPDSGVQVKGDFATDESNTSYFHDNGDKVYGGFFTTGNNAWYADNNGNLVKGPQ	1405
Db	1310	LYFRA--NGVQVKGEFVTDYGRISYYDNGSGDQIRNFRVRNAQQGFYFDNNGYAV----	1364
QY	1406	EIDGKWYHFDVETGQAKGAALVNGOOLYFDVDSGIVQVKGDFVTDGQGNYSYYDVNSGDK	1465
Db	1365	-----TGARTINGQLYFRA--NGVQVKGEFVTDHGRISYYDNGSGDQ	1406
QY	1466	KVNGFPTTGNNAWYADGQGNLAKGRKSIDNQDLYPDPATGKQVKGQLYSID--GRNYYPD	1524
Db	1407	IRNFRVRNAQQGFYFDNNGYAVTGARTINGQLYF--RANGVQVKGEFVTDYGRISYYD	1465
QY	1525	SCSGNMAK 1532	
Db	1466	ANSGERVR 1473	

RESULT 6

T30858
Glucosyltransferase - Streptococcus salivarius
C:Species: Streptococcus salivarius
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
R:Accession: T30858
R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.
Infect. Immun. 63, 609-621, 1995
A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for pri
A:Reference number: Z20909; PMID:95122197; PMID:7822030
A:Accession: T30858
A:Status: preliminary;
A:Molecule type: DNA
A:Residues: 1-1577 <SIM>

```
Db 951 PSQYTRVIAQNAKLPKEMGITSFEPAPQYVSQDGTFLDSIIENGAYAPEDRYDIAMSKN 1010
QY 909 NKYGSDDDLANALKALHANGIOAIADWVPDQIYQLPGEVVTAKRINSYGNPTFFDAYINN 968
Db 1011 NKYGSLKDLMDALRALHAEGISAIADWVPDQIYINLPGEVVTASRTNSYTPRPNAEYIN 1070
QY 969 ALYATNWKSGSDYQAOYGAFIDELKAKYPDMFTVMISTGKPIDPSTKIKOWEAKYFN 1028
Db 1071 SLYAAKTRTFGNDFOGKYGGAFIDELKAKYPAIFERVQISNGRKLATTNEKITQWSAKYFN 1130
QY 1029 GTNVLKGAGYVLSDDATGKYFTVNEGDFLPASFTGDQNAKTFPYVDGTGMAYYSTGN 1088
Db 1131 GSNIGTGARYVLODNATNQYFSVKAQOTFLPKQMT--EITGSGFRVRGDDVQYLSIGGY 1188
QY 1089 KAVNSIYEGGH--YYFDDKDGHMVTGSYKAEDGNDYFPLNGIQMDALYQDAQGSYYV 1147
Db 1189 LAKNTFIQVANGWYFDFKNGNMTGE-QVIDGKYFFLDNGLQLRHVLRQSGDGHVYY 1247
QY 1148 GRTGI-LYKGDNMWPFVDPNNANKTVFRYFDANNVMAIGYRNMYGOTYYFDE-NGFQAKG 1205
Db 1248 DPKGVQAFNG--FYDPAGP---RODVRYFDGNGQMYRGLHDMYGTTFYFDEKGTGIAKD 1301
QY 1206 QLLT-DDKGTHTFDEDNAGAKMKFVNVGDD--WYMDGNGNAVKQYPPNNQILLYNPE 1262
Db 1302 KTRFADGRTYFIPDTGNLAVNRPAQNPENKAWYLLDSNGYAVTGLQINGKQYYFDNE 1361
QY 1263 TGQVQKQFITDAQRTSYVDANSALKSSGFRPTNGSDWYYAENGYYVYKGFQVAENQD 1322
Db 1362 -GRQVKGHVT--INNORYFLDGSGEIAPSRFTEN-----N 1396
QY 1323 QWYFDDQTTGKAQGAADVDRLYFNPDSGVQVKGDFATDESNTSFYHGDNGDKRVVGG 1382
Db 1397 KWWYVD-GNGKLVKGAQVINGNHYYFNNDYS-QVKGAWA-----NGRYYDGSQGVASNQ 1449
QY 1383 FTTGNNWYADNNGNLVKGQEIQDKWYHDEVTGQQAAGALVNGQOLYFDVDSGTQ 1442
Db 1450 FIQIAANOWAYLNQDGHKVTGLQINNKKVYFG-----SNGAQ 1487
QY 1443 VKGDFVTDGQNTSYVDVNSGDKVNGPFTTGDNAWYADGQNLAKGRKSIDNQDLYPD 1502
Db 1488 VKGKLLT-VQGKCFDAHTGQVNVRFVEARGCWYFNSAGQAVTGGQVNGKQLYPD 1546
QY 1503 PATGKQVQKQVLSIDGRNYFSDSGSNMAKNR 1534
Db 1547 -GSGRQVKGRYVVGSKRFLCDAKTGELQRR 1577

RESULT 7
T30857
Glucosyltransferase - Streptococcus salivarius
C:Species: Streptococcus salivarius
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30857
R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.
Infect. Immun. 63, 609-621, 1995
A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for pri
A:Reference number: Z20909; MUID:95122197; PMID:7822030
A:Accession: T30857
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1449 <STM>
A:Cross-references: UNIPROT:Q55264; UNIPARC:UPI00000B166B; EMBL:L35495; NID:G662378; PID
C:Genetics:
A:Gene: gtfL

Query Match 39.6%; Score 3264; DB 2; Length 1449;
Best Local Similarity 44.3%; Pred. No. 6.7e-139;
Matches 729; Conservative 197; Mismatches 423; Indels 296; Gaps 47;

QY 1 MEKKLHYKHLKVKHWTIAVA--SIGLVSUVAGT-----VSAED----- 39
Db 1 MDKKVHYKMHYKQKWVTIATVGLSLGAVSAVSLGTGVDGVQADEHTDATVAIPDITVDT 60
```

```
QY 40 -KVANDTTA--QATGV-----DTGQDOAT---TNDANTNTTDT---DTADQSANTNQDOA 86
Db 61 GTVSNDDTTAAQPTTAAVAAATNDVATDOATPTATFDLTDTTNTVAANAADVTVAATVGTDR 120
QY 87 GDSQSNQOQAKQDPTANTDRNQADNSQTNNOQATDOATSPATDGTGVOR----- 135
Db 121 A---TTNDTTATNDTAVDTTN--NNTTTDTTVDRAATTERRATGARRGTGGRATPV 175
QY 136 ----RDAANVAT-----AADQEGTAPSEQEKSAALSIDNVKLDGKYYVQADGSYK 184
Db 176 NGNTNNWNTVTVVNDLPAATNNVVDGFS-----HIKTINGQYYVEDDGTIR 224
QY 185 KNFAITVNGOMLYFSDDTGALSSTSYSPSQ--GTTNLVDDFSSH-----NKAYDSTA 235
Db 225 KNYVLIERIGSGYFNAETGELSNOKEYRFDKNGGTGSSADSTNTVTVNGDKNAFYGTTD 284
QY 236 KSFELVNGYLTANSYRYPAGILRNGQWTSASNENDLRPVLMSWMPKQTOQAVYVNTMKY 295
Db 285 KDIELVDGYFTANTWYRPKEILKDGKEWTAENDKRPLLTVMPWPSKAIQASYLNTMKEQ 344
QY 296 -LSANETEVTNETSQVDLANKAOSLOTKEQKITSDNSTQMLRTAMEAFVAAQPKNMST 354
Db 345 GLGTNQT-YTSPSSQTMQDAALEVQKRIEERIAREGNTDMLRTTIKNFVKTPQGNST 403
QY 355 ENFNKGDHLQGGALLYTNSDLTPWANSDYRLNLRPTQODG--TKKYFTGEGGEGYEFLL 412
Db 404 ENLDNNDHLQGGALLYNNDSRTSHANSDYRLNLRPTSTQKHNPKYTKDTSNGGEFLL 463
QY 413 SNDVNSNPVQAEQLNQLHLYLNNQDVIWMDKDNFQGVRVDAVDNADLLQVYSNYF 472
Db 464 ANDIDNSNPAVQAEQLNQLHLYLNNQDVIWMDKDNFQGVRVDAVDNADLLQIASDYF 523
QY 473 KDNVKTUSEANALAHISITLEAWSLNDNQYNDTNGTALSIDNSSRLTSLAVLTQPGOR 532
Db 524 KAKYGADQSDQAIKHLSTLEAWSHNDAYNEDTGAQLPMDPMLHALVYSLLRPIGNR 593
QY 533 IDLSNLISESVNKERANDTAYGDTIPTYSFVRAHDSQVTVIAKIVKEIDTNSDGYTPT 592
Db 584 SGVEPLISLNSL-DRSESGSKRMANYAFVRAHDSQVSIQIQLKNEINPQSTGNTT 642
QY 593 LDQLKDAFKIYNEDMAKNVYTHYNIIPAAYALLSNMESVPRVYVYDLYTDGQYMAKK 652
Db 643 LDEMKAFAEINYKDMRSANKQYQYNIQYSAVALMTHKDTVPRVYVYDLYTDGQYMAKK 702
QY 653 SPYDAIATMLOGRYAYVSGGSEEVHKVN--CNGQ-----ILSSVRYQDLMSADDTG 705
Db 703 SPYDAIETLLKGRIRYAGGGQDMKNYIGYNTNGWDAAAGVLTSVRYGTGANSASDT-G 761
QY 706 TDLRSRTGLVTLVSNDPNLDLGGDSLTVNMGRAHQAAYRPLILGTQGVQSYLKSDSTN 765
Db 762 TAETRNQMAVIVSNQPALRL--TSNLTINMGAAHNRQAAYRPLLLTNDGVATYLNDSAN 820
QY 766 -IVKYTDANGNLTFADDIKYVSYVDMGSLAVWVPVGAQGDQVRAAATNQKADGKSL 824
Db 821 GIVKYTDGNGNLTFSANBIRGNPQVGYLAVWVPVGAENQDVRVAFSKEKNSSGLVY 880
QY 825 KTSALDSQVIYEGFSNFQDPFANDADYTNKKIAENADFPFKLGTITSPMAQVYSATDG 884
Db 881 ESNAALDSQVIYEGFSNFQDPFVQNPSTQYTNKKIAENANLFSKSWGITSFEPQYVSSDDG 940
QY 885 SPLDSIIQNGYAFSDRYDLAMSNNKKNYKSKDDLANALKALHANGIOAIADWVPDQIYQLP 944
Db 941 SPLDSVIQNGYAFSDRYDIHGSNNKKNYKSKLADLKAALKSLHANGIOAIADWVPDQIYQLP 1000
QY 945 GBEVVYAKRTNSYGNPTFDAYINNALYATNTKSSGSDYQAOYGGAFDLDELKAKYPMFTV 1004
Db 1001 GBEVVYATRVNNGYGETKGAIIIDHSLYAAKTRTFGNDYGGYKGAFLDELKELYPQIFDR 1060
QY 1005 NMISTGKPIDSTKIKOWEAKYFNGTNVLGKAGVYVLSDDATGKTYFTVNEGDFLP---- 1060
Db 1061 VQISTGKMTTDEKITQWSAKYTNMGNTNIDRGSYVNLKNGLNG--YGTNGGKVSPLPKVVG 1119
```


QY 1061 ---ASFTGD--ONAKTGFYYDGTG-----MAYYSTSGNAKVNSTRIVEG-GHYYYFDKDG 1108
DB 1120 SNQSTNGDNQNG-----DGSCKFEKRLFSVRYYNNQYAKNAFTKQNDGNVYFDSNG 1173
QY 1109 HMVTSYKAEADGNDYYFLPENGIMRDALYQDAGNSYYGRTGILYKGDWYFPVDPNNA 1168
DB 1174 RMAVGE-KTIDGQYFFLANGVQLRD----- 1198
QY 1169 NKTVFRYFDANNVMAIGYR-NMYGQTYYPDENG-FOAKGOLLTDDKGTGHYFEDNGAMAK 1226
DB 1199 -----GYRQNRGQVYYDQNGVLNANGK--QDPKP-----DNNNNASGR 1236
QY 1227 NKFNVNGDD-WYMYDGNAGNAVKQYFPVNNQILYFNPETGVQVKGPITDAQGRTSYYDAN 1285
DB 1237 NQPVQIGNVWYADYDNGKRVCHQNINGOELFFD--NNGVQVKGRVTNE-NGAIRYYDAN 1294
QY 1286 SGALKSGSFFTPNGSDWYVAENGYYVYKPKQVAENQDOWYFDTQTGKQAKGAAGVDGRD 1345
DB 1295 SG-----EMARNR----- 1302
QY 1346 LYFNPDSGVQVKGDFATDSGNSTSFYHGDNGDKVKGFFTTGNNAWYADNNGNLVKGPQ 1405
DB 1303 -----FAEIEPG-----VMAFYFNNDGTAVKGSQ 1325
QY 1406 EIDGKWHYFDEVTGQQAAGNAALVNGQOLYFDVDSGLQVKGDFVTDQGNSTSYDVNSGDK 1465
DB 1326 NI-----NGQDLYFD--QNGRQVKG--ALANVDGNLRYDVNSGSL 1362
QY 1466 KVNQFFTTGDNWYADGQNLAKGRKSIDNODLYDPDPAATGQVKGQOLYSIDGRNYYFDS 1525
DB 1363 YRNRFEHI--DGSWYFYDGNVAKGMVNINGQNLFFD--NNGKQIKGHLVRVGVRYFDP 1420
QY 1526 GSGNMAKNRFRIG-DOWIFYGNDG 1549
DB 1421 NSGEMAVNRWVEVSPGMWYVFDGEG 1445

RESULT 8
T30552
glucosyltransferase N - Streptococcus salivarius (fragment)
A:Species: Streptococcus salivarius
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30552
R:Jaffe, R.I.
submitted to the EMBL Data Library, February 1998
A:Description: Streptococcus salivarius V1477 gtfN.
A:Reference number: Z20854
A:Accession: T30552
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1449 <JAF>
A:Cross-references: UNIPROT:O68542; UNIPARC:UPI00000B10FD; EMBL:AF049609; NID:G29335545;
C:Genetics:
A:Gene: gtfN

Query Match 39.4%; Score 3243; DB 2; Length 1449;
Best Local Similarity 44.1%; Pred. No. 5.8e-138;
Matches 726; Conservative 198; Mismatches 425; Indels 296; Gaps 47;

QY 1 MEKKLHYKLHKVKKHVVTVAVA--SIGLVLVAGT---VSAED----- 39
DB 1 MDKKVHYKHKKVKKQWTVATVGLSLGAVSLSLGTNDGVVQADEHTDATVAIPDITVDT 60

QY 40 -KVANDTTA--QATGV-----DTGQDQAT---TNDANTNTDT---DTADQSANNTQDQA 86
DB 61 GTVSDNTTAAQDPTTAAATNDVATDQATPTATFDLTDTTNTVAANA VDTVATVCTDRA 120

QY 87 GSDQSNQDQAKDPTANTDRNQADNSQTDNNQATDQATSPATDGTSVQR----- 135
DB 121 A---TTNDTATNDTAVDTTN--NNTTTDTTNTNRAATERRATGARRGPTGGRBATPV 175

QY 136 ----RDAANVAT-----AADQEGQTAPSEQBSAALSILDNVKLLIDGKYVYVQADGSYK 184

DB 176 NGNTNANNVTVVVNDLPAITNNVVTGDS-----HIKTINGQYVYVEDDGTJR 224
QY 185 KNFAITVNGQMLYFSDTGCALSSSTYSFSQ--GTTNLVDDPSSH-----NKAYDSTA 235
DB 225 KNYVLIRIGGSQYFNAETGELSNQKEYRFDKNGGTGSSADSTNTNTVNGDKNAFYGTGTD 284
QY 236 KSFEVLNGLVYLTANSYRPAIILNGQWTWEASNENDLRPVLMSWPDKQTOVAVVYNNKY 295
DB 285 KDIELVDGTFYANTWYRPEILKQKWEWTASTENDKRPLLTVMVFSKAIQASLYLTKQ 344
QY 296 -LSANETEVNETSVDLNKEAQSIQTKEQKITSNDSTQWLRTAMEAFVAAQPKNMST 354
DB 345 GLGTWQT--YTSFSSQTMQDAALEVQKRIEGRIAREGNTDLRTTIKNFVKTPQGNST 403
QY 355 ENFNKGDHIOGALLYTNSDLPWANSYRLLNRTPTQODG--TKKYFTGEGGEGYEFLL 412
DB 404 ENLDVNDHLOGGALLYNNDSRTSHANSYRLLNRTPTSQTGKHNPKYTKDTSNGGFEFL 463
QY 413 SNDVNSNPVQAEQLNOLHLYLMMWGDIVMGDKDANFDGVRVADVNVNADLLQVYSNYP 472
DB 464 ANDIDNSPAPVQAEQLNMLHYIMNIGTTGSEDEFNFDGVRVADVNVNADLLQIASDIF 523
QY 473 KQNYKVTDSANALAHISILEAWSLNDNQYNETDNGTALSIDNSRLTSLAVLTKPQGOR 532
DB 524 KAKYGADQSDQAIKHLSSILEAWSHNDAYNEDTKGAQLPMDPMHLALVYSLRPIGR 583
QY 533 IDLSNLISESVNKERANDTAYGDTIPTYSFVRAHDSVQTVTAKEIKIDNSDGYTTF 592
DB 584 SGVEPLISLSNL--DRSESGKSKRMANYAFVRAHDSVQSIIGQIKNEINPQSTGNTT 642
QY 593 LDOLKDAFKYNEDMAKNTKTYHNI PAAYALLSNMESVPRVYVGLYTDGQYMAKK 652
DB 643 LDEMKAFEIYNKDMRSANKQTYQNI PSAYALMTHKTDVTPRVYIGMDITDDGQYMAOK 702
QY 653 SPYYDAIATMLQRIAYVSGGQSEEVHKVN--GNQ-----ILSSVRYGQDLMASDDTQ 705
DB 703 SPYYDAIETLLKGRIRYAAAGQDMKVNIIGYNTNGWDAAGVLTSVRYGTGANSADT-G 761
QY 706 TDLSTSGLVTLVSDPNLDLGGDSITVMGHAHNAQAYRPLILGTQGVQSVYKDSDTN 765
DB 762 TAETRNQGMVIVSNQPALRL--TSNLTINMGAHRNQAYRPLLLTTNDGCVATYLNDSAN 820
QY 766 -LVKYTDANGNLTTFTADDIKGYSTVDMSGYLAVWVPVGAQGDGVRAADTQKADGKSL 824
DB 821 GLVKYTDGNGNLTFSANEIRGNPQVGYLAWVPVGAQGDGVRAADTQKADGKSL 880
QY 825 KTSAAALDSQVIYEGFSNFQDFANNDADYTNKKIAENADFFKLGITGSFEMAFQYVSATDG 884
DB 881 ESNAALDSQVIYEGFSNFQDFVPNPSQYTNKKIAENANLFSKWGITSFEFAFQYVSSDDG 940
QY 885 SPLDSIIQNGYAFSDRYDLAMSKNNKYKSKDDLAKALHANGQAIADWVPDQIYQLP 944
DB 941 SPLDSVIQNGYAFTRYDIDGMSKDNKYGSLADLKAALSHAVGISAIADWVPDQIYQLP 1000
QY 945 GEEVTVAKRTNSYGNPTFDAYINNALYATNTKSSGSDYQAOYGGAFDLDELKAKYPMFTV 1004
DB 1001 GEEVTVATRVNNGYGETKCAIIDHSLYAAKTFTFNDYQCKYGGAFDLDELKRLYPOIFDR 1060
QY 1005 NMISTGKPIDPSTKIKOMEAKYFNGTNVLGKAGYVLSDDATGKYFTVNEENGDFLP---- 1060
DB 1061 VQISTGKMTTDEKITSKAKYMGNTNILDGSEYVLKNGLNG--YGTNGGKVS LPKVWG 1119
QY 1061 --ASTGTD--ONAKTGFYYDGTG-----MAYYSTSGNAKVNSTRIVEG-GHYYYFDKDG 1108
DB 1120 SNQSTNGDNQNG-----DGSCKFEKRLFSVRYYNNQYAKNAFTKQNDGNVYFDSNG 1173
QY 1109 HMVTSYKAEADGNDYYFLPENGIMRDALYQDAGNSYYGRTGILYKGDWYFPVDPNNA 1168
DB 1174 RMAVGE-KTIDGQYFFLANGVQLRD----- 1198
QY 1169 NKTVFRYFDANNVMAIGYR-NMYGQTYYPDENG-FOAKGOLLTDDKGTGHYFEDNGAMAK 1226
DB 1199 -----GYRQNRGQVYYDQNGVLNANGK--QDPKP-----DNNNNTSGR 1236

A:Gene: dsrA
C:keywords: glycosyltransferase; hexosyltransferase
F:78-870/Domain: catalytic #status predicted <CAT>
F:922-1290/Domain: glucan-binding #status predicted <GCB>

Query Match 34.0%; Score 2799; DB 2; Length 1290;
Best Local Similarity 44.3%; Pred. No. 4.4e-118; Indels 118; Gaps 34;
Matches 614; Conservative 201; Mismatches 454;

QY 164 LBNVLIDGKYVVOADGYSKKNFAITVNGQMLYFDSGTALSSSTYSFSOGTNNLVD 223
DB 1 MFNVKHVDGVFYGDGQPKKNFTTIIDGKPYFDKGTGALSND---KQVSELFSI 56

QY 224 FSSHKAYDSTAKSFELVNGYITANSWYPAIGLRNGQTWEASNENLRLPVLMSPWPKD 283
DB 57 GNKNNAVNTSSDNFTQLEGLHTASSWYRPKDIILKNRKPWAPSTVTDFRPLLMAMPDKS 116

QY 284 TOVAVVNYMK--YLSANETEVNTSVDLNLKEAQSIQTKTEOKITSDNSTQWLRTAME 341
DB 117 TQVYLYNWKDQGLSGTHHFSNDENMRT-L7AAMQAQVNIETKIGQLGNTDMLKTAMT 175

QY 342 AFVAQPKKWNMTSTENFKG--DHLOGGALLYTNSDLTPWANSDYRLNLRTPTOQDG--TKK 398
DB 176 QYIDAQPNWIDSE--AKGDDHLQGGALLYTNSDMSPKANSDYRKLSTRTPKNQKQIADK 233

QY 399 YTEGEGEGYFLLSNDVNSNPVQAEQNLQHLNWKWDIVMGDKDANFDGVRDAVD 458
DB 234 Y----KQGGFELLANDVNSNPVQAEQNLNHLNMMNIGSILQNDQANFDGYRDAVD 289

QY 459 NYNADLLOVSNFYKDYKVTSEANALAHISILEAWSLNDQYNEDTNGTA--LSIDN 515
DB 290 NYVADLLQAGEYAKAAGV--DKWTRERINIOFWKTGEMKIQMSKHMATSKLSMDF 346

QY 516 SSRLTSLAVLTQPGQORIDLSNLISSEVNKERANDTAYGDTIPTTYSFVRHDSVQTVIA 575
DB 347 PLHLAIKVALNPNDKSGLEPTREHSLVK-RITDDKENVAQPNYSFIRAHDSVQTVIA 405

QY 576 KIVKEKIDNSGY--TFTLDQLKAPKLYNEDMAKNTYTHYNIIPAAVALLSNMESV 633
DB 406 DIIDKINPASTGLDSTVTLDOIQAFDIYNADLKAQVYTPYNIIPASYALLNPKDTI 465

QY 634 PRVYGDLTDDGYNMAKSPYDAITATMLQRIAVVSGQSEEVHVKVNGNQIILSSVRY 693
DB 466 PRVYGDMDTDDGYNMAKSPYQALDALKARIYAAGGQTKMKNYFPDQSGVMTSVRY 525

QY 694 GDILASADTQGTDLSTSGVTLVNSDNLIDGG--DSLTVNMGRAHANQAYRPLITLTK 752
DB 526 KGKAMTASD--GNOETRYOGIGLVNRRPDLKSLDKDEKMDMGAHKNQDYRPVLLTTK 584

QY 753 DGVSYLKDSQDNI VKYITDANGNLFTADDI KGYSTVDMSGYLAVVVPVGAQDQDVRA 812
DB 585 SGLKVTSTANAPVVR--TDANGQLTFKADWYGVNDPQVSGYIAAWVPVGAQENQDARTK 643

QY 813 ADTNOKADGKSLKTSALDSOVITYEGFSNFPQANNDADYTNKKAENADFFKGLGITSF 872
DB 644 SETTQSTGVSVHNSAALDSQVITYEGFSNFPQFPPTTDFEFTNIKAQNVNLFKDMGITSF 703

QY 873 ENAPQVVSATDGSFLDSIIQNGYAFSDRYDLAMSKNKYGSKDDLANALKALHANGIOAI 932
DB 704 ENAPQVRSASDRSFLDAIVQNGYAFDTRYDIGVNTPTKYGTADNLLDALRALHGGIOAI 763

QY 933 ADWVPDQIYQLGEEVVTAKRNSYGNPTFDAYNNALYATNTKSSGSYQAOYGAFLD 992
DB 764 NDWVPDQIYNLDEQVLTAIRTDGSDHTYGSVIDHTLYASKTVAGGI-YQQOYGGAFLE 822

QY 993 ELKAKYPDMFTVMNISTGPIDPSTKI KOWEAKYFNGTNVLKGGYVLSDDATKYFTV 1052
DB 823 QUKTQIPQLPQOKQISTDQPMNPDIQIKSWEAKYFNQSGNIQGRGAWVILKDMGTQYFNV 882

QY 1053 NENGDFLPASFTGDQNAKTGFYYDGTGMAYYSTSGNKAVNSFTIEGHHYVTKDGHMYT 1112
DB 883 SDAQTFLPQLLGEK-AKTGFVTRGKETSFYSTSGYQAQSAFICDNGNMYYPDDKGMVV 941

QY 1113 GSYKAEDGNDYVFLPNGIQMRDAIYQDAQNSYYIGRTGILYKGDWNPVFPVDPNNANKTV 1172
DB 942 GN-QVINGINIFLPLNGIELQDA-YLVHDGMYYYNNIG-KQLHNTYYQDKOKN-----992

QY 1173 FRYFDANNVMAIGYRNMYGQ-----TYTFDENGFOAKGQLLTDGK-THYFDEDNAMAK 1226
DB 993 PHYFEDGHMAOGIIVTIQSDDGTPVQYFDENGKQKQGVAVKSGDHLHYFDGASGNMLP 1052

QY 1227 NKFVNVGD-DWYMDGNNAVKGYPPVNNQILYFNPETGVQVKGFITDAQRTSYIDAN 1285
DB 1053 KSWGRLAGDSWLYVDKGNVATGKOTINNQTQYFEN-DGROQIKNFKELADG-----1103

QY 1286 SGALKSSGFFTPNGSDWYEAENGYYVKGFKQVAENQDQWYFDQTTGKQAKAAKYDGRD 1345
DB 1104 -----SWLYLNN-----KQVA-----VTGEQI-----INGOT 1125

QY 1346 LYFNPDSGVQVKGFATDESNGTSFYHGDNGDKVYGGVFTTGNNAWYADNNGNLVKGFQ 1405
DB 1126 LYFGND-GRQFKGTTTHINATGESRYDPPDSGNMITDRPERVGDNQWY-----1172

QY 1406 EIDGKWHYFDEVTGQOAKGAALVNGQQLYFDVDSGIQVKGDFVTGQGNSTSYDVNSGDK 1465
DB 1173 -----FGVD--GVAVTGDRIIKGNLYFN-QNGIQMKGHLRLE-NGIMRYYDADTGL 1221

QY 1466 KVNGFPTTGDNAWYADGQGNLAKGRKSIDNQDLYFDPATGKQVKQLVSIIDGRNYPDS 1525
DB 1222 VNRNREVLSDSGSWYFGQGVPTGVQVINGQTLTFD-ADGRQVKQQRVIGNQRYWMDK 1280

QY 1526 GSGNMAK 1532
DB 1281 DSGEMK 1287

RESULT 13
B97033
uncharacterized protein, related to enterotoxins of other Clostridiales [imported] -Clu
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004
C:Accession: B97033
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clu
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: B97033
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2817 <KUR>
A:Cross-references: UNIPROT:O97K42; UNIPARC:UPI00000CA0A0; GB:AE001437; PIDN:AAK79053.1
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC1079

Query Match 7.9%; Score 654.5; DB 2; Length 2817;
Best Local Similarity 20.4%; Pred. No. 3.3e-21;
Matches 438; Conservative 271; Mismatches 660; Indels 775; Gaps 114;

QY 34 TVSAEDKVANDTTAQTGVDVTGQDQAT-----TNDANTTTTDTTADQSAANTNQDQAGSD 89
DB 130 TPKNESSINNEA--PIIPKDTSKTSSTSSAQTKGSDNNNIPSNNTTSTNTSKNENPSNTD 186

QY 90 QSNQDQAKQDNTANDRQADN-----SQTNNQATDQATSPATDGTSTQVRDAANVAT 143
DB 187 IKTTAPANAPTKIDTPNNQSDSALAKNALSNNLAAD---SSQTSKVTSSNNDAKPVNT 243

QY 144 AADQEGQTAPEQESKSAALSNDNLKIDG-----KYYVQADSGYKKNFAITVNGOML 196
DB 244 T-----STDKKASNLNDSQ---DGVWTKDGKKYYV--NGVQKQGFQ-SINKSIY 288

QY 197 YFDSB-----TGALSSTSTYSFSGQTTNLVDFDSSHKNKAYDSTAKSFELVNGYLTANS---249
DB 289 YFNDGDSMGTGWLKNSNSYYFDASGVMLTGLQNINGTYGFNDGKLLTGLQAINNNY 348

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 18:57:35 ; Search time 209.688 Seconds
(without alignments)
5228.676 Million cell updates/sec

Title: US-10-797-821-38
Perfect score: 8237
Sequence: 1 MEKKLHYKLHKVKKHWTTIA.....FVRIGDQWYFGNDGAATNL 1554

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot 05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8237	100.0	1554	2	Q8KZL5_9STRE
2	4375	53.1	1512	2	Q9WXJ5_9STRE
3	4360.5	52.9	1506	2	Q56CX8_9STRE
4	4014	48.7	1599	2	Q00599_STRSL
5	3818	46.4	1518	2	Q00600_STRSL
6	3711.5	45.1	1577	2	Q54178_STRGN
7	3672.5	44.6	1575	2	Q9LCH3_STROR
8	3660	44.4	1462	1	GTFD_STRMU
9	3502.5	42.5	1590	2	Q59983_9STRE
10	3485	42.3	1476	1	GTFB_STRMU
11	3481.5	42.3	1590	2	Q55263_9STRE
12	3481	42.3	1597	1	GTFI_STRDO
13	3472.5	42.2	1592	1	GTFZ_STRDO
14	3429.5	41.6	1577	2	Q55265_STRSL
15	3317.5	40.3	1455	1	GTFC_STRMU
16	3264	39.6	1449	2	Q55264_STRSL
17	3243	39.4	1449	2	Q68542_STRSL
18	3242	39.4	1454	2	Q68A94_LEUME
19	3142	38.1	2835	2	Q89Q02_LEUME
20	3080.5	37.4	1365	1	GTF5_STRDO
21	3047	37.0	1338	2	Q9WXJ4_9STRE
22	3024	36.7	1522	2	Q6TXV4_LEUME
23	3017	36.6	1527	2	Q9ZAR4_LEUME
24	3015	36.6	1527	2	Q8KRE1_LEUME
25	2964.5	36.0	1508	2	Q52224_LEUME
26	2959.5	35.9	1508	2	Q9EZH5_LEUME
27	2958	35.9	1477	2	Q91466_LEUME
28	2922	35.5	1561	2	Q5SBM8_9LACO
29	2912	35.4	1330	2	Q84CN4_LEUME
30	2821	34.2	1595	2	Q5SBM3_LACSK
31	2802	34.0	1290	2	Q48756_LEUME

32 2643 32.1 1463 2 Q5SBM6_LACFE
33 2422 29.4 2057 2 Q9RE05_LEUME
34 2306 28.0 1016 2 Q9LCJ7_LEUME
35 2285.5 27.7 1772 2 Q5BN3_LACRE
36 2285.5 27.7 1781 2 Q4JCS4_LACRE
37 2285 27.7 1781 2 Q5SBL9_LACRE
38 2280.5 27.7 1781 2 Q4JLC7_LACRE
39 2220.5 27.0 1772 2 Q5SBN0_LACRE
40 1409 17.1 1619 2 Q5SBN0_LACRE
41 1384 16.8 1231 2 Q5SBN1_LACRE
42 1379 16.7 591 2 Q8VUH3_STRMU
43 1076 13.1 522 2 Q8VV10_STRSA
44 670 8.1 374 2 Q6ZXI9_9LACO
45 654.5 7.9 2817 2 Q97K42_CLOAB

Q5sbm6 lactobacill
Q9re05 leuconostoc
Q9lcj7 leuconostoc
Q5bn3 lactobacill
Q4jcs4 lactobacill
Q5sbl9 lactobacill
Q4jlc7 lactobacill
Q5sbn0 lactobacill
Q5sbn1 lactobacill
Q8vuh3 streptococc
Q8vv10 streptococc
Q6zx19 lactobacill
Q97k42 clostridium

ALIGNMENTS

RESULT 1
Q8KZL5_9STRE PRELIMINARY; PRT; 1554 AA.
AC Q8KZL5_9STRE PRELIMINARY; PRT; 1554 AA.
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Glucosyltransferase.
GN Name=gtfu;
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1310;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21958684; PubMed=11960691; DOI=10.1016/S0304-4165(01)00240-9;
RA Hanada N., Fukushima K., Nomura Y., Sempuku H., Hayakawa M.,
RA Mukasa H., Shiroza T., Abiko Y.;
RT "Cloning and nucleotide sequence analysis of the Streptococcus
RT sobrinus gtfu gene that produces a highly branched water-soluble
RT glucan.";
RL Biochim. Biophys. Acta 1570:75-79(2002).
DR EMBL; AB089436; BAC07265.1; -; Genomic_DNA.
DR HSSP; P06653; 1H8G.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW binding_1; 5.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1554 AA; 171676 MW; 6981BCC1DAE24A73 CRC64;

Query Match 100.0%; Score 8237; DB 2; Length 1554;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1554; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEKKLHYKLHKVKKHWTTIAVASIGLVSIVAGTVAEDKVDNDTTAQATGVDTGDDQA 60
DB 1 MEKKLHYKLHKVKKHWTTIAVASIGLVSIVAGTVAEDKVDNDTTAQATGVDTGDDQA 60
QY 61 TTNDANTNTTDTTADQSANNTQDQAGSDQSNQDQADTANTDRNQADNSQTDNNQAT 120
DB 61 TTNDANTNTTDTTADQSANNTQDQAGSDQSNQDQADTANTDRNQADNSQTDNNQAT 120
QY 121 DOATSPATDGTSVQRDAANVATAADQGTAPSPQEKSAALSLDNVKLIDGKYTVQAD 180
DB 121 DOATSPATDGTSVQRDAANVATAADQGTAPSPQEKSAALSLDNVKLIDGKYTVQAD 180
QY 181 GSYKKNFAITVNGQWLYFSDSDTSGALSTSTYSFSGTTLNVDLDFSSHKNKAYDSTAKSPFL 240
DB 181 GSYKKNFAITVNGQWLYFSDSDTSGALSTSTYSFSGTTLNVDLDFSSHKNKAYDSTAKSPFL 240
QY 241 VNGYLTANSWVRPAGILRNGQTWEASNDLRPVLMSWNPDKDTQVAYVYNNKYLISANE 300

Db 241 VNGYLTANSWYRPAGLIRNGQTWEASNENDLRPLVMSWPDKDTQVAYVNYMKNLSANE 300
QY 301 TEVNTSETVDLNKEAQSQTKEIQKITSNDSTQWLRTAMEAFVAAQPKWNMSTENFNKG 360
Db 301 TEVNTSETVDLNKEAQSQTKEIQKITSNDSTQWLRTAMEAFVAAQPKWNMSTENFNKG 360
QY 361 DHLQGALLYTNISDLTPWANSYRLLNRPFTQDGTGKFTTEGEGEGYFLLSNVDNSN 420
Db 361 DHLQGALLYTNISDLTPWANSYRLLNRPFTQDGTGKFTTEGEGEGYFLLSNVDNSN 420
QY 421 PVVQAEQLNQLHYLNMWGDVMDGDANFDGVRVDAVDNADLLQVYNYFKDNTKVTD 480
Db 421 PVVQAEQLNQLHYLNMWGDVMDGDANFDGVRVDAVDNADLLQVYNYFKDNTKVTD 480
QY 481 SEANALAHISILEAWSINDQYNEDTNGTALSIDSNSRLTSLAVLTQPKQGRIDLSNLIS 540
Db 481 SEANALAHISILEAWSINDQYNEDTNGTALSIDSNSRLTSLAVLTQPKQGRIDLSNLIS 540
QY 541 ESNVKERANDTAYGDTIPTYVSFVRAHDSVQTVIAKIVKEKIDTNSDGYTFTLDQLKDAF 600
Db 541 ESNVKERANDTAYGDTIPTYVSFVRAHDSVQTVIAKIVKEKIDTNSDGYTFTLDQLKDAF 600
QY 601 KIYNEDMAKVNKTYTHYNIIPAAVALLSNMESVPRVYVYGDLYTDDGQYMAKSPYYDATA 660
Db 601 KIYNEDMAKVNKTYTHYNIIPAAVALLSNMESVPRVYVYGDLYTDDGQYMAKSPYYDATA 660
QY 661 TMLQGRIAVYSGQSEBEVHKVNGNNOILSSVRYGQDLSADDTQGTDLRTSLGLVLSN 720
Db 661 TMLQGRIAVYSGQSEBEVHKVNGNNOILSSVRYGQDLSADDTQGTDLRTSLGLVLSN 720
QY 721 DPNLDLGGSLTVNMGRAHANQAYRPLILGTGDKGVQSYLKSDSTNIYKVYTDANGNLTFPA 780
Db 721 DPNLDLGGSLTVNMGRAHANQAYRPLILGTGDKGVQSYLKSDSTNIYKVYTDANGNLTFPA 780
QY 781 DDILKGYSTVDMGYLAVVPVGAQGDVRAADTNQKADGKSLKTSAAALDSQVIEYEGS 840
Db 781 DDILKGYSTVDMGYLAVVPVGAQGDVRAADTNQKADGKSLKTSAAALDSQVIEYEGS 840
QY 841 NFQDFANNADYTNKKIAENADPFKKLGITSEFMAPQYVSATDGSPLDSIIQNGYAFSDR 900
Db 841 NFQDFANNADYTNKKIAENADPFKKLGITSEFMAPQYVSATDGSPLDSIIQNGYAFSDR 900
QY 901 YDLAMSKNNKYGSKDDLALNALKALHANGIOAIADWVPDQIYQLPGBEVTVAKRTNSYGNP 960
Db 901 YDLAMSKNNKYGSKDDLALNALKALHANGIOAIADWVPDQIYQLPGBEVTVAKRTNSYGNP 960
QY 961 TPDAYINNALYATNTKSSGSDYOAOYGGAFDLDELKAKYPDMFTVNMISTGKPIDPSTKIK 1020
Db 961 TPDAYINNALYATNTKSSGSDYOAOYGGAFDLDELKAKYPDMFTVNMISTGKPIDPSTKIK 1020
QY 1021 QWEAKYFNGTNLVKGAGYVLSDDATGKFTVNEGDFLPASFTGQNAKTGPYYDGTGM 1080
Db 1021 QWEAKYFNGTNLVKGAGYVLSDDATGKFTVNEGDFLPASFTGQNAKTGPYYDGTGM 1080
QY 1081 AYYSTSGNKAVNSFIYEGGHYYFDKDHMTGYSYKAEDGNDYFPLNGIQMRDAIYQDA 1140
Db 1081 AYYSTSGNKAVNSFIYEGGHYYFDKDHMTGYSYKAEDGNDYFPLNGIQMRDAIYQDA 1140
QY 1141 QGNSYYTGRGILYKGDNWPVVDPPNNANKTVPRYFDANNVMAIGYRNMVGGTYYPDENG 1200
Db 1141 QGNSYYTGRGILYKGDNWPVVDPPNNANKTVPRYFDANNVMAIGYRNMVGGTYYPDENG 1200
QY 1201 FQAKGQLLTDKGTHTYFDEDNAGMAKNKFVNVGDDWYMDGNAGVKGQYPVNNQILYFN 1260
Db 1201 FQAKGQLLTDKGTHTYFDEDNAGMAKNKFVNVGDDWYMDGNAGVKGQYPVNNQILYFN 1260
QY 1261 PETGVQVKGQFTIDAQGRYSYDANSALKSSGFFTPNGSDWYTYAENGYYVYKFKQVAEN 1320
Db 1261 PETGVQVKGQFTIDAQGRYSYDANSALKSSGFFTPNGSDWYTYAENGYYVYKFKQVAEN 1320
QY 1321 QDQWYYFDQTTGQAKGAACVDRDLYFNPDSGVQVKGDFATDSEGNSTFYHGDNGDKVY 1380

Db 1321 QDQWYYFDQTTGQAKGAACVDRDLYFNPDSGVQVKGDFATDSEGNSTFYHGDNGDKVY 1380
QY 1381 GGFPTTGNNAMYADNNGNLVKGFOEIDGKWKYHFDVETGQQAAGALVNGQQLYFDVDSG 1440
Db 1381 GGFPTTGNNAMYADNNGNLVKGFOEIDGKWKYHFDVETGQQAAGALVNGQQLYFDVDSG 1440
QY 1441 IQVKGDFVTDGOGNTSYDVNSGDKKVGFFTTGNNAMYADGQGNLAKGRKSIDNQDLY 1500
Db 1441 IQVKGDFVTDGOGNTSYDVNSGDKKVGFFTTGNNAMYADGQGNLAKGRKSIDNQDLY 1500
QY 1501 FDPATGKQVKQVQLVSDIGRNNYFDSGSGNMAKNRFRVIRGQDWIYFGNDGAATNL 1554
Db 1501 FDPATGKQVKQVQLVSDIGRNNYFDSGSGNMAKNRFRVIRGQDWIYFGNDGAATNL 1554
RESULT 2
Q9WXJ5_9STRE PRELIMINARY; PRT; 1512 AA.
AC Q9WXJ5; DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE GTF-S.
GN Name=glTF;
OS Streptococcus criceti.
OG Plasmid pAMI.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1333;
RN [1]_TaxID=1333;
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=HS-6;
RA Inoue M., Fukui K., Miyagi A.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB026123; BAA77237.1; -; Genomic_DNA.
DR HSP; P06654; IMPE.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 2.-
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Plasmid.
SQ SEQUENCE 1512 AA; 167145 MW; 4C03D9C8C601FC14 CRC64;
Query Match 53.1%; Score 4375; DB 2; Length 1512;
Best Local Similarity 55.6%; Pred. No. 1.6e-182;
Matches 882; Conservative 188; Mismatches 393; Indels 122; Gaps 27;
QY 1 MEKKLHYKLHKYKHHVTTIAVASIGLVSILVGAG---TVSAEDKVANDTTAQTATVGDVG 56
Db 1 MEKKLHYKLHKYKHHVTTIAVASIGLVSILVGAG---TVSAEDKVANDTTAQTATVGDVG 56
QY 57 QDQATTNDANTTTTDTDTA---DQSANTNQDQAGSDQSNNOQAKQDQTANTDRNQADNSQ 113
Db 57 QDQATTNDANTTTTDTDTA---DQSANTNQDQAGSDQSNNOQAKQDQTANTDRNQADNSQ 113
QY 53 -EQNTSASQNEVDSDAQAATDAKATSEQAASVSDTINSE-----TQVQNVQGV 102
Db 53 -EQNTSASQNEVDSDAQAATDAKATSEQAASVSDTINSE-----TQVQNVQGV 102
QY 114 TDNQ-----ATDQATSPATDGTSGVRRDAANVATADQ-----EGQTAP-----SEQK 159
Db 114 TDNQ-----ATDQATSPATDGTSGVRRDAANVATADQ-----EGQTAP-----SEQK 159
QY 103 SANNQPEAPAAPQA--ASNNNTNTATSEANTNTATSEAAAPAAENRTAEKADLSQBEA 160
Db 103 SANNQPEAPAAPQA--ASNNNTNTATSEANTNTATSEAAAPAAENRTAEKADLSQBEA 160
QY 160 AALSIDNKLIDGKYVYQADGSKYKNFAITVNGQMLYFSDTGTGALSSTTSFSOGTTN 219
Db 160 AALSIDNKLIDGKYVYQADGSKYKNFAITVNGQMLYFSDTGTGALSSTTSFSOGTTN 219
QY 161 AALSIDNKLIDGKYVYQADGSKYKNFAITVNGQMLYFSDTGTGALSSTTSFSOGTTN 220
Db 161 AALSIDNKLIDGKYVYQADGSKYKNFAITVNGQMLYFSDTGTGALSSTTSFSOGTTN 220
QY 220 LVDDFSSHNKAYDSTAKSPELVNGYLTANSWYRPAGLIRNGQTWEASNENDLRPLVMSW 279
Db 220 LVDDFSSHNKAYDSTAKSPELVNGYLTANSWYRPAGLIRNGQTWEASNENDLRPLVMSW 279
QY 221 LVSDSFINNKAEDSRAKSFELVDGYLTAESWYRPTKILENGKTWVDSKETDLRPVLTSM 280
Db 221 LVSDSFINNKAEDSRAKSFELVDGYLTAESWYRPTKILENGKTWVDSKETDLRPVLTSM 280
QY 280 PKDQTVAYVNTMKNYLSANETEVNTNETSQVDLNKEAQSIOQIKOKITSDNSTQWLRTA 339
Db 280 PKDQTVAYVNTMKNYLSANETEVNTNETSQVDLNKEAQSIOQIKOKITSDNSTQWLRTA 339
QY 281 PKDQTVAYVNTMKNYLSANETEVNTNETSQVDLNKEAQSIOQIKOKITSDNSTQWLRTA 339
Db 281 PKDQTVAYVNTMKNYLSANETEVNTNETSQVDLNKEAQSIOQIKOKITSDNSTQWLRTA 339
QY 340 MEAFVAAQPKWNMSTENFNKGDHQLGGALLYTNISDLTPWANSYRLLNRPFTQDGTGK 399
Db 340 MEAFVAAQPKWNMSTENFNKGDHQLGGALLYTNISDLTPWANSYRLLNRPFTQDGTGK 399

Db 340 MAAFVATQGRWNSDSQFQKNDHLOGGALLYTNKLTWEADSKYRLNLRTPTRQDGKTHY 399
QY 400 FTREGGEGYEFLLSNDVNSNPVQAEOLNOLHLMNWGDI VMGDKDANFDGVRVDAVN 459
Db 400 SKADEYGGYEFLLANDVNSNPVQAEMLNQIHYLMNWSI VMGDKDANFDGIRVDAVN 459
QY 460 VNADLLQVSNYPKONYKVTDSANALAHISILEAWSLNDNQYNEDTNGTALSIDNSSRL 519
Db 460 VDADTJQLYTNFVNAVYGVDSQAQALAHISILEAWSYNDNYNQDTNGAALAMDNGRL 519
QY 520 TSLAVLTQPGORI-DLSNLI-SESUNKERANDTAYGDTIPTYSFVRANDSEVQTVIAKI 577
Db 520 SLLYTLTRPLSERTPGSLTLIKSEYGLTDRTKDKYGDTPQSVFVRAHDSEVQTVIAQI 579
QY 578 VKEIDITNSDGYFTFLDQLDKAFKIYNEDMAKVNKYTHYNIPAAVALLSNESPRVY 637
Db 580 IREKIDPTDGTFTFLDQLKQAFDIYNKDNWSVEKHYTHYNIPAAVAMLSNESVTRY 639
QY 638 YGDLYTDDGOYMAKSPYDAIATMLQGRIAVYSGQSEVHK----- 680
Db 640 YGDLFTDDGOYMETKSPYDAINTLLRARIRYAAGGQTWE-HKAYTPSAAMKAKNPDSGS 698
QY 681 VGNNOILSSVRYGQDLMSADDTQGTDLSTRTSLVLSNDPNLDL-GGDSLTVMNGRAH 739
Db 699 VLGNSVLYSVRFQDVMGADDMTGGLAKTGMFSLIANPPELELDANEELKVNVGKTH 758
QY 740 ANQAYRPLLTGKGVQSVYKDSNTNIVKYTDANGNLTFADDIKGYSTVDMSGYLAVWV 799
Db 759 ACQAYRPLLTGKGLQKYLNSDNTNLTVAQDGFITFKGSEIKGYKOVGVNGYLSVWV 818
QY 800 PVGAKDGQDVRVAADTNQKADG-KSLKTSAAALDSQVIYEGFSNFQDFANNADYTNKKTIA 858
Db 819 PVGAKSDQDIRVAASATKANAKGDKSVTASQALDSQLIYEGFSNFQDFVQKDAQYTNKKTIA 878
QY 859 ENADFFKGLGITSFEMAPQYVSATGSLDLSIIQNGYAFSDRYDLAMSKNNKYGSKDILA 918
Db 879 ENTDLFKWGVTSFEMAPQYVSATGSLDLSIIQNGYAFSDRYDLAMSKNNKYGSKDILA 938
QY 919 NALKALHANGIOAIAADWPQIYQLPGREVVTAKTNSVGNFTFPAYINNALYANTKSS 978
Db 939 NALKALHANGIOAIAADWPQIYQLPGREVVTAKTNSVGNFTFPAYINNALYANTKSS 998
QY 979 GSDYQAQYGGAFDLBKARYPDMFTVMNIMSTGKPIDPSTKIKQWEAKYFNGTNVLGKAG 1038
Db 999 GKDFQAKYGGEFELAELOKQYPEMFTAKMISTGKPIDSSVKLKEWSAQYFNGTNVLGRGTD 1058
QY 1039 YVLSDDATGKYFTVNEGDFLPASTGDNQAKTGYPYDGTGMAYYSTSGNKAVNSFIYEG 1098
Db 1059 YVLSDEGTGKYFTVNEKEGDFLPVLTGDKAETGFPYNDGKGMTYFTTAGSQAQSDFTVA 1118
QY 1099 GHYYFDFKDHMTGSKYAEADNDYFELPNGIQMRDAIYQDAQNSYVYGRGILYKG-- 1156
Db 1119 GNTYFDFYTGHWVTP-NGINTKFYFELPNGVMLKDAVMEDDRGVSYYGKTGMVYKGRS 1177
QY 1157 -DNWYFPVDPNNANKTVFRYFDANNVMAIGYRNMTGQTYFDPENGFOAKGOLLTDKGT-T 1214
Db 1178 NNEWFAMTD--SKGQLRFRHFDNYGPMVGLVTHIGNVQYDYDEEGFOVKGFVTDKAGQT 1235
QY 1215 HYFEDNGAMAKNFVNVGDDWYVMDGNGNAVKGOYPVNNQILYFPNPTGVQVKGQFID 1274
Db 1236 RYFDKNTGNLVKGQPFQNGHWHYSDQCLIAKGAQTIKGOKLYFPAKTGAQVKGDFVTD 1295
QY 1275 AQGRTSYDANSALKSSGFFTPNGSDWYYA-ENGYYVYKGFQVAVENQDQWYFQDTGK 1333
Db 1296 KDGNTFFYSGDGTDLAVSTFFSTGNNWAFYADENGH----- 1332
QY 1334 OAKGAAYDGRDLYFNPDSGVQVKGDFATDESGNTSFYHGDNGDKVVGGEFTT--GNNAW 1391
Db 1333 -AKGEKTINGOKLYFDTKTGOAKGRFVRDAGK-IRPFDADTGLVUTNSFLETKAGSNQW 1390
QY 1392 YYADNNGNLVKGQFBDGKWHYFDEVTCQAKGAALVNGQQLYFDVDSGIOVKGDFVTDG 1451

Db 1391 YYMGADGYAVRGHQTITQSRHMYFDAETQQAQKIV-----VTDA 1429
QY 1452 QGNTSYDYDNGDKKYNVGFFTTGDNAWYADGOGNLAQGRKSIDNQDLYFDPATGKQVKG 1511
Db 1430 NERKIFYDANTGDRVNVNQVLV-NGSWYFFGYDGAATVGRDIRGOHLYFNP-DGTOAKG 1487
QY 1512 QLVISIDGRNYYPDSGSGNMAKQRFV 1536
Db 1488 TTVKIDNRIYTFDADSGELTSVRYI 1512
RESULT 3
Q56CX8 98TRE PRELIMINARY; PRT; 1506 AA.
AC Q56CX8;
DT 10-MAY-2005 (TREMBLrel. 30, Created)
DT 10-MAY-2005 (TREMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TREMBLrel. 30, Last annotation update)
DE Glucosyltransferase-T.
GN Name=gtft;
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
RN NCBI_TaxID=1310;
RX [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=B13N;
RA Kuwahara N.;
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY966490; AAX76986.1; -; Genomic_DNA.
KW Transferase.
SQ SEQUENCE 1506 AA; 167016 MW; 94B882EF2C17C451 CRC64;

Query Match 52.9%; Score 4360.5; DB 2; Length 1506;
Best Local Similarity 54.7%; Pred. No. 6.7e-182;
Matches 869; Conservative 202; Mismatches 349; Indels 169; Gaps 25;
QY 1 MEKLLHYKLHKVKKHWVTIAVASIGLVSLVGAG---TVSADKQVANDTTA-QATVGVDT 55
Db 1 MERKLLHYKLHKVKKHWVTIAVASAGLASIVGAGSLQTVSADDLAKEQAASQQAANQ 60
QY 56 GQDQATTNDANT---NTTDTDTADQSANQDQAGSDQSNQDQAK-QDTANTDRNQADN 111
Db 61 NEDEVASDAADTASAKATSEKVVQSSDTNSE---TNQVETKQASAKESADAVAKAQAPQ 117
QY 112 SQTNNQATQATSPATDGTSVQRDAANVATA--ADQEGTAPSEQEKSAALSIDNVKL 169
Db 118 A-----GPAIT-SQVASESSVAPSKAEADKAAAGSVSQNEEALSLANIKK 164
QY 170 IDGKYVYVOADGSKYKKNFAITVNGOMLYFSDTGSSTSTVSQGTTLNVDDFSHNK 229
Db 155 IDGKYVYVNAOGSKYKKNFAITVCGOMLYFSDTGSSTSTVSQGTTLNVDDFSVNNK 224
QY 230 AYDSTAKSFELVNGYLTANSWYRPAIGILRNQGTWEASNENDLRPVLMSWMPDKDTQVAVY 289
Db 225 AFDSEKSFELVDGVLTAESWYRPAKILLENKGTWVDSKETDLRPVLMSWMPKDTQVAVL 284
QY 290 NYMKYLSANETEVNTSETSQVDLNKBAOSIQTKIEQKITSNOSTOWLTAMEAFVAAQPK 349
Db 285 NYMSKALGGKE-EFTTETSQTLNTAAELIOAKTEARVSKEGQTKWLREMAAFVATQSR 343
QY 350 WNMSTENFNKGDHLOGGALLYTNSDLTTPWANSYVRLNRTPTQODGTKKYFTEGEGGVE 409
Db 344 WNKQSEYDKADHLOGGALLYTNNNLTWANSNRLNRTPTQODGTKKYFTEGEGGVE 403
QY 410 FLLSNDVNSNPVQAEOLNOLHLMNWGDI VMGDKDANFDGVRVDAVDNVDNADLLQVYS 469
Db 404 FLLANDVNSNPVQAEMLNQIHYLMNWSI VMGDKDANFDGIRVDAVDNVDNADLLQVYT 463
QY 470 NYFKDNYKVTOSEANALAHISILEAWSLNDNQYNEDTNGTALSIDNSSRLTSLAVLTQKP 529
Db 464 NYFNSVYGVNKSEQAQALAHISVLEAWSYNDNYNQDTNGAALAMDNGRLFSLLYTLTRPI 523

```
QY 530 GORI-DLSNLI-SESUNKERANDTAYGDTIPTYSVRADHSEVOTVIKIVKEKIDTNSD 587
Db 524 NERTPGMSTLIRKSEYGLTDRTKNDKYGDTQPSYFVRADHSEVQTVIAQIIKEKIDPTFD 583
QY 588 GYTFILDQLDKAFKYNEDMAKVNKYTHYNTIPAAVALLLSNMESVPRVYVYGLVTDQDQ 647
Db 584 GYTFILDQLDKAFKYNEDMAKVNKYTHYNTIPAAVALLLSNMESVTRVYVYGLVTDQDQ 643
QY 648 YNAKSPYYDAIATMLQGRYAVSGQSEE-----VHKVNGNQNQLGSV 691
Db 644 YNAKSPYYDAINTLLRARIRVAAAGQIMEHNSYKPSAAMKAHPDAGNVLGNSEVLNSV 703
QY 692 RYGODLMSADDTQGTDLSTSLGLTVLSNDPLDLG-GDSLTVNMGRAHQAQYRPLILG 750
Db 704 RFQGDVMSADDTGGKLAKTSGMFTLISNNPELADVNBEIKVNVGKIHAGQAYRPLLT 763
QY 751 TKDQVQSYLKSDSTNIVKYTDANGNLTFTADDIKGYSTVDMGSLVAVVWVPGAKQDQVR 810
Db 764 TKDGLQKYLNSDSTKLTKADKDGFTFKGSEIKGYQVEVNGYLSVWVPGAKAQDIR 823
QY 811 VAADTNQKAD-GKSLKTSAAALDSQVIYEGFSNFPQFANNADYTNKKIAENADFPKKG 869
Db 824 VAPSTAAGKEKARTYASQALESQLIYEGFSNFPQFVKQDSQYTNKKIAENTDLFKAMGV 883
QY 870 TSFEMAPQVVSATDGSFLDSIIIONGYAPSDRVDLAMSNNKNGSKDDLALNALKHANGI 929
Db 884 TSFEMAPQVVSATDGSFLDSIIIONGYAPSDRVDLAMSNNKNGSKDDLALNALKHANGI 943
QY 930 QAIADWVPQIYQLPQEEVVTAKRTNSYGNPTFDAYINNALYATNTKSSGSDYQAYGGA 989
Db 944 QAIADWVPQIYQLPQKEVVTASRVNDYGRVKVDQPLVEKLYLANTKSSGKDFQAYGGE 1003
QY 990 FLDELKAKYPMFTNMISTGRPIDPSTKIKOWEAKYFNGTNVLGKAGYVLSDDATGY 1049
Db 1004 FLAELQKYPFMTMTKSTIDPSVKLKEWSAKYFNGTNVLDRTDYLSDGTYGK 1063
QY 1050 FTVNENGDFLPASFTGDQNAKTFYDGTGMAIYSTGNKAVNSFTYEGHYYFDKDH 1109
Db 1064 FTVNEKGDFLPASLTGNKDAKTFYDGTGMAIYSTGNKAVNSFTYEGHYYFDYTH 1123
QY 1110 MYTGSYKABDGDNDYFPLPNGIOWRDAIYODACNSVYVYGTGILYKG--DN-WYFVDPN 1166
Db 1124 MYTGP-NVINTKFEYFPLPNGIMLKDAIKQDEKRSVYVYGTGILYKG--DN-WYFVDPN 1182
QY 1167 NANKTVRYFDANNVMAIGYRNMYGQTYFDFENGFOAKGQLTDDKG-THYFDENGAWA 1225
Db 1183 GQWR--FRHFDYRGFMSIGLVTINQNVYDENGFOVKGFEFTDODGQTRYFDQSGNLV 1240
QY 1226 KMKFVNVDGDMYMDGNGNAVKQYPPVNNQILYFNPETGVQVKGQFIDTAQGRTSYYDAN 1285
Db 1241 KGQFLNKGDNWYILD----- 1255
QY 1286 SGALKSSGFFTPNGSDWYVAENGYYVYKGFQVAENQDQWYFDDTGTGQAKGAAGVDGRD 1345
Db 1256 -----DQ--GLVAKGAQITKGK 1271
QY 1346 LYFNPDSGVQVKGDFATDSGNTSFYHGDNGKWKVGFFTTGNNAWYADNNGNLVKGPQ 1405
Db 1272 LYFDTKTGVQVKGDFVTDKDGNTFFYSGDTGDLILGQFFSTGNNAWYADENGHV----- 1326
QY 1406 EIDGKWHYFDEVTGQAKAALVNGQOLYFDVDSGLQVKGDFVTDQGNSTSYDVNSGDK 1465
Db 1327 -----AKGAKTIRGQKLYFDTKTGOQAKGRFIRDDKG-VRYYDADTGL 1369
QY 1466 KVNQFFTT--GDNWYVYADQGNLAKGRKSIDNQDLYFDPATGKQVKGQVLS-IDGRNYY 1522
Db 1370 VTNAFLETAGSNQWYMGADGVAVKGNQTIKNQHYFPAETGQQAAGLIIVTDANGKXF 1429
QY 1523 FDSGSGNMAKNRFRIGDQWYFVNDGAA 1551
Db 1430 YDTFTGSRVNVQFVLVNGWNYFFGYDAA 1458
```

```
RESULT 4
Q00599_STRSL
ID Q00599 STRSL PRELIMINARY; PRT; 1599 AA.
AC Q00599;
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DT 01-FEB-2005 (TREMELrel. 29, Last annotation update)
DE Glucosyltransferase S.
GN Name=gtfK;
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Lactobacilliales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 25975;
RX MEDLINE=93381463; PubMed=8371114;
RA Giffard P.M., Allen D.M., Milward C.P., Jacques N.A.;
RT "Sequence of the gtfK gene of Streptococcus salivarius ATCC 25975 and evolution of the gtf genes of oral streptococci.";
RL J. Gen. Microbiol. 139:1511-1522(1993).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 25975;
RX MEDLINE=92148377; PubMed=1838391;
RA Giffard P.M., Simpson C.L., Milward C.P., Jacques N.A.;
RT "Molecular characterization of a cluster of at least tow glucosyltransferase genes in Streptococcus salivarius ATCC 25975.";
RL J. Gen. Microbiol. 137:2577-2593(1991).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 25975;
RA Jacques N.;
RT Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
RN [4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 25975;
RA Giffard P.M., O'Connor D.M., Milward C.P., Simpson C.L., Jacques N.A.;
RT "Sequence of the gtfK gene of Streptococcus salivarius ATCC 25975 and the evolution of the gtf genes of oral streptococci.";
RL Mol. Microbiol. 0:0-0(1992).
DR EMBL; Z11872; CAA77898.1; -; Genomic DNA.
DR EMBL; Z11873; CAA77901.1; -; Genomic DNA.
DR FIR; S22737; S22737.
DR HSP; P06653; 1H8G.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 3.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1599 AA; 176480 MW; 24B77869E152B707 CRC64;

Query Match 48.7%; Score 4014; DB 2; Length 1599;
Best Local Similarity 51.7%; Pred. No. 9.8e-167;
Matches 851; Conservative 215; Mismatches 430; Indels 150; Gaps 40;

QY 1 MEKLYKHLKHKHWTIIVASIGLSLVGAGTVSA-----EDKVANDTTAQATVGDV 54
Db 1 MENKRYKHLKHKQVMTLAVASVALATIVGGSVATSSLASAEETNNSGSPSTTTVGEN 60
QY 55 TG-----QQATNTDANT-NTTDDP-----TAPQSANTN-QDQAGSDQSNQDQAKQDTAN 103
Db 61 TNPVVEKEVGTTEVANTSNTATTERRAEVTADKPAGTTVQPNSGT-----T 106
QY 104 TDRNQA--DNSQTDNNQATDQATSPATDGTSPQRDAANVATAA--DQEQQATPSEOEKA 160
Db 107 SDRAAAVEAEKPEPTAKPEVATKPEATTS---EVAANAGVAAPTTEKSKLSEAEIKA 163
QY 161 ALSLDNVLK-IDGKYVYVQADGSKYKKNFAITVNGQMLYFDSGTGALSSSTYSFSQGTTN 219
Db 164 AVSLDNIKKDKGYLLLEDGSHKKNFAITVNGQVLYFD-ENGALSSSTYSFSQGTTN 222
```


Db 61 LVTSPEATKADRTNKADVLTPAKETNAVETAT-TTNTQATAEAAATTATADVA-VA 118
QY 113 QTDNNQATQOATSPATDGTSGVQRDA--ANVATAADQOQOAPSPQESKAALSLSNVKL 169
Db 119 AVPNKEAVVTDAPATTEKABEQPATVKAENVNTEVKAPAEALKDSEVEAALSLSNIKN 178
QY 170 IDGKYYVQADGSGYKKNFAITVNGQMLYFSDTSGALSTSTYSFSGOTTNLVDFFSSHNK 229
Db 179 IDGKYYVNVNEDSGHKNFAITVNGQLLYFKGD-GALTSSTSYSTFTPGTTNIVDGFSINNR 237
QY 230 AYDSTAKSRELNGYLITANSYRPAIGLRNGOTWEASNENDLRPVLMSWMPDKDOTOVAV 289
Db 238 AYDSSEASPELIDGYLTADSWYRPAIIKDGVTWQASTAEDFRPLLMWPNVDTGVNVL 297
QY 290 NYMKNYLSANETEVNTEISQVLDLNKEAQSIQTKIEQKITSDNSTOWLRTAMBAFVAAOPK 349
Db 298 NYMSKVFNL-DAKYSSTDQOETLKVAAKDIOIKIEQIOAKSTQWLRETIISAFVKTPQ 356
QY 350 WNMSTENFNKG---DHLQGGALLYTNSDLTPWANSYRLLNRTPTQODGT--KKYPTGG 404
Db 357 WNKETENYSGGGEDHLOGGALLYVNDSTRTPWANSYRRLNRTATNQTGTIDKSLDEQS 416
QY 405 E---CGYEFLLSNVDNSPVVQAEOLNOLHYLMNWGDIUMGDKDANFDGVRVDAVDNV 460
Db 417 DPNHMGDFLLANDVDLSNPVVQAEOLNOLHYLMNWGSIUMGDKDANFDGIRVDAVDNV 476
QY 461 NADLLQVSNFYKNDYKVTDSANALAHISILEAWSLNDNQYNEDTNGTALSIDNSSRLT 520
Db 477 DADMLQLYTNYPREYGVNKSANALAHISVLEAWSLNDHNDYNDKTDGAALAMENKORLA 536
QY 521 SLAVLTK-----QPGQRIDLSNLISESVNKERAND-----T 551
Db 537 LLFSLAKPIKERTPAVSPLYNNTFNTRQDEKTDWINKDGSKAYNEDGTVKQSTIGKYNE 596
QY 552 AYGDTPPTYSFVRAHDESVQTVIAKIVKEKIDTNSDGYTFLDQLDKAFKINVEDMAKN 611
Db 597 KYGDASGNVYFIRAHONNVQDIIAEIIRKEINPKSDGFTITDAEMQOAFEINYKMWLSSD 656
QY 612 KTYTHYNI PAAYALLSNMESVPRVYVYGDLYTDDGOYMAKSPYDAIATMLQRIAYVS 671
Db 657 KKYTLNNI PAAYAVMLQNETITRVYVYGDLYTDDGHYMETKSPYDITVNLMSRLKYVS 716
QY 672 GGQSEVH-----KVGNNQILSSVRYGQDLMSADDTQGTDLRSRTSGLVTLV 718
Db 717 GGQQRSYWLPDGRKMDNSDELRYRTNEVYTSVRYGKDINTANDTEGSKYSRTSGQVTLV 776
QY 719 SNDPNLDLGDGS-LTVNMGRAHANOAYRPLILGTQDGVQSYLKQSD--TNTVXYTANG 774
Db 777 ANNPKNLDQSAKLNEMGKIHANQYRALIVGTADGINKFTSDADAIAGVYVYKETSNG 836
QY 775 NLTFTTDDIKGYSTVDMSGYLAVVVPVGAKGDQVRVAADTNTQKADGK-SLKTSAALDSQ 833
Db 837 VLTFFGANDIKGYETFDMSGFVAWVPVGASDNQD IRVAPSTEAKKGELTLKATEAYDSQ 896
QY 834 VYIEGFSNFQDF--ANNADYTNKKIAENADFPKKLIGITSFEMAPYVSATGCSFLDLSII 891
Db 897 LIYIEGFSNFQTTIDGSDPSVYTNKIAENVDLPFKSMGVTSFEMAPQFVSADDDGTFDLSVI 956
QY 892 QNGYAFSDRYDLAMSNNKYKSKDDLANALKALHANGIOAIADWVPDQIYOLPGEVWTA 951
Db 957 QNGYAFADRYDLAMSNNKYKSKEDLRDALKALHAGIOAIADWVPDQIYOLPGEVWTA 1016
QY 952 KRTNSYGNPTFDAYINNALYATNTKSSGSDYOAOYCGAFDELKAKYPMFTVMMLSTGK 1011
Db 1017 TRTDGAGRKIADAIIDHSLYVANSKSGKDYQAKYGEFEFLAELKAKYPENFKVMMLSTGK 1076
QY 1012 PIDPSTYKIQWEAKYPNGTNVLGKAGYVLSDDATGKYFTTVNENGDFLPASFTGDQNAKT 1071
Db 1077 PIDDSVKLQWKWAEYFNGTNVLERGVYVLSDEATGKYFTVTKEGNFIPQLTGTGKEKIT 1136
QY 1072 GFYYDGTGMAYYSTGNKAVNSFIYEGGHYYTFDKDGHMVTGYSKAEADGNDYY-FLPNGI 1130

Db 1137 GPSSDGKIGITYPCTSGTOAKSAFVTFNGNTYYTFDARGHMVWTNSEYSPNGKDVYRFLPNCI 1196
QY 1131 QMRDAIYQDAQNSYYYGRGTGILYKGDNMWYPPVDPNNAWK-----TVFRYFDANNVMA 1183
Db 1197 MLSNAFYIDANGNTYLYNSKGQYKGG--YTKFDVSETDKDGKESKVVKFRYFTNEGVA 1254
QY 1184 IGRNMYGOTYYFDENGFOAKQLLTDKGTHTYFDEDNGAMAKNFVNVDGDDWYMDGNG 1243
Db 1255 KGVTVIDGTQYFGSDGFOAKDLVTFKGTYYTFDAHTNGTNGITWRNNGKWWYFEDANG 1314
QY 1244 NAVKGOYPVNNQILYFNPETGVQVKGQFITDAQGRTSYVDANSALKSSGFFTPNGSDWY 1303
Db 1315 VAATGAQVINGQKLFPN-EDGSQVKGGVVKNADGTSYKYKEGFGELVTNEFTTIDGNVWY 1373
QY 1304 YAENGYYVYKGFQKVAENQDQWYYPDQTTGKQAKGAKVGDGRDLYFNPDPGQVKGDFATD 1363
Db 1374 YA-----GANGKTVTGAQVINGQHLFYFNAD--GSQVKGGVVKN 1409
QY 1364 ESGNTSFYHGDGDKVVGGFPTTGNNAWYADNNGNLVKGFOEIDGKWTHFDEVTGQOAK 1423
Db 1410 ADGTYSKYNASTGERLTNEFTTIDGNWYIYGANGKSVTGEVKI-----1453
QY 1424 GAALVNGQOOLYFVDSGIGQVKGDFVTDGQGNTSYYDVNSGDKKXNGFFTTGDNAWYYADG 1483
Db 1454 -----GDDTYFAKDGQVKGQTVSAGNGRISYYYGDSGRAVSTWIEIQGVYVYFDK 1507
QY 1484 QG 1485
Db 1508 NG 1509

RESULT 6

Q54178 STRGN PRELIMINARY; PRT; 1577 AA.
AC Q54178; Q54247;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Glucosyltransferase.
GN Name=gtfG;
OS Streptococcus gordonii.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX [1] NCBI_TaxID=1302;
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Challis;
RX MEDLINE=96157084; PubMed=8586195;
RA Vickerman M.M., Sulavik M.C., Clewell D.B.;
RT "Molecular analysis of Streptococcus gordonii glucosyltransferase
phase variants.";
RL Dev. Biol. Stand. 85:309-314(1995).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Challis;
RA Minick P., Vickerman M.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; U12643; AAC43483.1; -, Genomic_DNA.
DR PIR; B41898; B41898.
DR HSSP; P06653; 1HCX.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW binding_1; 5_
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1577 AA; 177805 MW; 5AE0328DC5B08D18 CRC64;
Query Match 45.1%; Score 3711.5; DB 2; Length 1577;
Best Local Similarity 48.0%; Pred. No. 1.6e-153;
Matches 788; Conservative 226; Mismatches 461; Indels 167; Gaps 38;

```
QY 1 MEKKHVKLVKHKHVVTTAVASIGLVSL-----VGAGTVSABD-----X 40
Db 2 MEKKVHYKMHKKNWVAIAVTSIALIVAPKALGLESVVIYADDAQVNVVKEQSAVQSK 61
QY 41 VANDTTA-----QAATGVDTGQDQATTN-DANTNTTDTTADQASANTNQ--- 83
Db 62 DSQTTSDKATDSQLEVKEQASSETTYQASATNPNTANEQTTQDQKEVTSRTDSRHE 121
QY 84 -DAQSDQSNNOQAQD-----TANTDRNQ-----DNSQTDNNQATDQATSPATDGT 131
Db 122 LTQKTBDSSSEKSGSQEPKVAQDAESTDKTQAALQAKQDSRANDQDEETENVAKATVSD 181
QY 132 SV-----QRRDAANVATAADQEGQAPSEQKSAALSQDNKVLIDGKYVYVQ 178
Db 182 KIIATPKKERLPBPAPQKESITEKMLAAQ-AQAPVNTTEHDDV-LAHIKTIDGKYVYVQ 239
QY 179 ADGSYKKNFATVNGOMLVFSDTGALSSSTSTYSFSQGTINLVDDFSHNKAVDSTAKSF 238
Db 240 DQGTVKKNFAVELNGKILYFDAQETGALVDSNEYFQOQGTSSLNNEFTQKNAFYGTGTDKI 299
QY 239 ELVNGYLTANSWYRPAGILRNQGTWEASNEENDLRPVLMSWPDQDTQVAVVNNVNYK--- 295
Db 300 ETVDGVLTAADSWYRPKFLKDGKTTASTETDLRPILLMAWPDKRTQINLYNMGQENLG 359
QY 296 LSNATEVETNETSQVDLNKEAQSIQTKIBOKITSDNSTOWLRTAMBAFAVAAQPKNMSTE 355
Db 360 IGAFE-----SKTEQVLLTNVAVQVQRKIBERISKEGDTKWLRTLMSAFVKTQPNWNKTE 415
QY 356 NFNKG---DHLOGGALLYTNSDLTPWANSQDYRLINRTPPTQODGTYKYPTEGEGGYEFLL 412
Db 416 SETTGKNDHLOGGALLYTNSDKTSHANSRYRLINRTPPTSQGTPKYFIDKSNNGGYEFLL 475
QY 413 SNDVNSNPVQAEQLNQLHYLNMWGDIVMGDKDANFQDVRVDAVDNVDNADLLQVYSNYF 472
Db 476 ANDFNSNPVQAEQLNQLHYLNMWFGSIVANDPTANFQDVRVDAVDNVDNADLLQIASDYF 535
QY 473 KONYKYTDSANALAHISILEAWSLNDQNYEDNTGTALSIDNSSRLTSLAVLTQKQGR 532
Db 536 KSRYKVGSEEEALKHLSILEAWSNDPDYNDKTKGAQLAIDNKLRLSLYSFMRKLSIR 595
QY 533 IDLSNLISSEVUNKERANDTAYGDTPTYSFVRAHDSQVQTVIAKIVKEKIDNDSQGYTPT 592
Db 596 SGVEPTITNSLN-DRSTENKNGERTANYIFVRAHDSQVQTVIADIRENINPNTDGLTPT 654
QY 593 LQQLKDAFKIYNEDMAKNKTYTHYNI PAAYALLSNMESVPRVYVYDLYTDDGQYMAKK 652
Db 655 MDELKQAFKIYNEDMRKADKKTQFNIPTAHALMLSNKDSITRVYVYDLYTDDGQYMEKK 714
QY 653 SPYDAIATMLQGRYAYVSGGQSEEVHKYN-----GNNQILSSVRYGQDLMGADDTQ 704
Db 715 SPYHDAIDALLRARIKYVAGGQDMKYTVYNGVPREADKWSYNGILTSVRYGTGANEATD-E 773
QY 705 GTDLSRTSLVTLVSNPDNLIG-GDSLTVNNGRAHANAQYRPLILGTGQVQSYLKQSD 763
Db 774 GTAETRTQGMAYIASNNPNLKLNEWDKLQVNNGAHKQYRVPVLLTTKDGISRYLTDBE 833
QY 764 T--NIVKYTDANGNLTFTADDIKGYSTVDMQSVGLAVWVPVGAQDQDVRVAAVTQKQAD 821
Db 834 VPQSLWKKTDANGIUTFDMDNLAGSYNVQSVGLAVWVPVGAQENQDARVTSKKNAGS 893
QY 822 KSLKTSAAALDSQVIEGFSNFQDFANNADYTNKTKIAENADFPKGLGITSFEMAPQYSA 881
Db 894 QYVESPALDSQILIEGFSNFQDFATRDQDYTNKYIAKNVNLFEKVGWTSFELPQYVSS 953
QY 882 TQGSFLDSIIQNGYAFSDRYDLAMSKNNKYGSKDILANALKAHANGIOAIADWPDQYI 941
Db 954 QDGTFLDSIIQNGYAFEDRYDMAMSKNNKYGSLDLLNALLHALHSVNIQAIADWPDQYI 1013
QY 942 QLPGEVVTAKETNSYGNPTFPAYINNALYATNTKSSGSDYQAOYCGGAFDELKAKYPM 1001
Db 1014 NLPFGVEVVTATRVNNYGYREGAEIKELLYVANSKNTGTDYQGYGGAFDELKAKYPEI 1073
QY 1002 FTVANMISTGKPIDPSTKIKQWEAKYFNGTIVNLKGGAGYVLSDDATGKYFTVNEGD-FIP 1060
```

```
Db 1074 FERVQISNGQKMTTDEKITKWSAKYFNGTILGRGAVYVYLKDWGSEKYL-S-NKNGETALP 1132
QY 1061 ASFTGQONAKTGYDGTGMAYYSTSGNKAVNSFIY-EGGHYYVYFDKDHMTVTSYKABD 1119
Db 1133 KQLV-NKEASTGVPKDTNGFKFYSTSGNAQKDTFIQDENGWYVYFDNQGLVLTGA-REID 1190
QY 1120 GNDYFLPNCIORDAIYQDAQNSYYVYGRG--ILYK-----GDNWYPPVDPNNANKTV 1172
Db 1191 GKQLYFMKNGVQLRDALQEDENGNOYYDKTGAKVLNRYVTSQGNW----- 1237
QY 1173 FRYEDANNVMAIGYRNMYGQTYFYFDENGFOAKQOLL-TDDKGYHYFDEEDNGAMAKNFVN 1231
Db 1238 -RYFDAKGVMAKGLVKIGDQVQVFDQNGYQVKGKVVRAKDGKLYRFDKDSGNAVIRFAQ 1296
QY 1232 VGD---DWYMDGNGNAVKQVYVNNQILYFNPETGVQVKGQFITDAQGRTSYVDANSQA 1288
Db 1297 -GDNPSDWTYFYGADGVALTGLOKIGQOQTLFYF-QDGKQVKGQVVMVLADKSIRYFDANSGE 1354
QY 1289 LKSSGFFTPNGSDWYVAENGYVYKQVAENQDOWYVYDQTTGKQAKGAAKVDGRDLYF 1348
Db 1355 MAVNKE-----AEGA-----KNEWYFDO--DCKAVTGLKTINNQLVYF 1391
QY 1349 NPDGQVQVKGDPATDESNTSFYHGDNGDKVVGFFTTGNNNAWYVYADNNGNLVKGFEID 1408
Db 1392 DQD-GKQVKGQVVTLADKSIRYFDANSGEVMAVKNKPAEGAKNEWYVFDQDKAVTGLQTI- 1449
QY 1409 GKWYHFEVTEGQAKGAALVNGQQLYFDVDSGLOVKGDPFTDQCGNTSYVDVNSGDKQVN 1468
Db 1450 -----NKQVLYFGQD-GKQVKGQVVTLADKSIRYFDANSGEVMAVN 1488
QY 1469 GFFTTGDNAWYVYADQGNLAKGRKSIDNODLYFDPATGQVKGQVLYSIDGRNYYFDSGSG 1528
Db 1489 KFAEGAKNEWYVFDQDKAVTGLKTINNQLVIFG-QDGKQVKGQVVTVDGAERYFDPKSG 1547
QY 1529 NMAKRFVRIGD-QWYIFGNDG 1549
Db 1548 DMVRNKWIRLEDGTWYFDRNG 1569
RESULT 7
Q9LCH3 STOR PRELIMINARY; PRT; 1575 AA.
ID Q9LCH3 STOR PRELIMINARY; PRT; 1575 AA.
AC Q9LCH3;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Glucosyltransferase.
GN Name=gtfr;
OS Streptococcus oralis.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1303;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC10557;
RX MEDLINE=202311779; PubMed=10768934;
RX DOI=10.1128/JAI.68.5.2475-2483.2000;
RA Fujiwara T., Hoshino T., Ooshima T., Sobue S., Hamada S.;
RT "Purification, characterization, and molecular analysis of the gene
RL encoding Glucosyltransferase from Streptococcus oralis.";
RL Infect Immun. 68:2475-2483(2000).
DR EMBL; AB025228; BAA95201.1; -; Genomic_DNA.
DR HSSP; P06653; IHCX.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 4.
DR Pfam; PF02324; Glyco_hydro_70; 1.
DR Transference.
SQ SEQUENCE 1575 AA; 176792 MW; 772A26E4D7C2E543 CRC64;
```


Query Match	44.6%;	Score 3672.5;	DB 2;	Length 1575;
Best Local Similarity	47.4%;	Pred. No. 7.9e-152;		
Matches	779;	Conservative 228;	Mismatches 472;	Indels 165; Gaps 36;
Qy	1	MEKKLHYKLVKCHWVTVIAVIGIV---	SLVG--ACTVSAED----	KVANDTTAQ--- 48
Db	2	MEKKLHYKLVKCHWVTVIAVIGIV---	SLVG--ACTVSAED----	KVANDTTAQ--- 48
Qy	49	-----ATVGDVTGQQAATTNDANTNT	TDTDQASANTNQOAGSQSNNOQAK-QDT	101
Db	62	PGQOTPAQAKIAEQEAEKVTADKVT	DDVAASEKPAKPAENTEATVQNAQEP	KADPT 121
Qy	102	--ANTDRQADNSQTDNNQATDQATSP	ATDGTSVQRDAAVATAADQEG-----	QTA--- 152
Db	122	KEASTEKAABAEBVKAANAATIPKTE	VAAD-----QNKQARPTTAQDQEGD	KREKTAVED 176
Qy	153	-----PSEQEKSAAALS	-----	LDNVKLIDGKYVV 177
Db	177	KIVANPKVAKKORLEPPGSKGAIABRM	VADQAQAPVNVADHDDVLSHIKTIDG	KNYV 236
Qy	178	QADGSYKKNFAITVNGOMLYFSDTGAL	SSSTYSFSGQTTNLVDDPSSHKNKAYD	STAKS 237
Db	237	QDDGTWKKNFAVELNGRILYFDAETG	ALVDSNEYQFOQGTSSLNNEFSQKNAP	YGTITDKD 296
Qy	238	FELVNGYLTANSWYRPAIILNGQITW	EASNEIDLRLVMSWHPDKDTQVAYV	YNNKY-L 296
Db	297	IETVDGYLTADSWYRPFILDKGTWTAS	TETDLRLPLMAWPKRTQINLYN	MNQQL 356
Qy	297	SANETEVNTEQSDVLNKAQSIQTKIEQ	ITSDNSTOWLRTAMEAFVAAQPKMN	STEN 356
Db	357	GAGAFE--NKVEQALLTGASQVQRKIE	KIEKIEGEGDKWLRLMGAFTQPNW	NIKTES 414
Qy	357	FNKG-J-DHQQGALLYTNSDLTPWANS	DYRLNLRPTTQDDGTKKYFTFEGG	GGYFELLS 413
Db	415	ETTGTKKDLHQGALLYTNNKSPHADS	KFLNLRPTTSGTKPKYFIDKSN	GGEFELLA 474
Qy	414	NDVDSNPVQAEOLNQLHLMNWDGI	VMGDKDAMFDGVRDADVNADLN	QVSNYFK 473
Db	475	NDFDSNPVQAEOLNQLHLMNPFSGI	VANDPTANFDGVRDADVNADLN	QVSNYFK 534
Qy	474	DNVKTDSANALAHISILEAWSLNDQ	NEDTNGTALSIDNSSRLTSLAVLT	KPQQR 533
Db	535	SRYKGESEEEAKHLSILEAWSLNDQ	NEDTNGTALSIDNSSRLTSLAVLT	KPQQR 594
Qy	534	DLNLSISESVNKERANDTAYGDTIPT	YSFVRAHDSVQTVIAKIVKEKID	TNSDGYTFL 593
Db	595	GVEPTITNSLN--DRSEKKNGERMANY	IFVRAHDSVQTVIADIRENINPTDGL	TFM 653
Qy	594	DQLKDAFKLYNEDMAKWNKYTHYNI	PAAYALLSNMESVPRVYGDLYTD	DGOYMAKS 653
Db	654	DELQKAFKLYNEDMRKADKKYQFNI	PTAHALMLSNKDSITRYVYGDLYTD	DGOYMEKS 713
Qy	654	PYDATTATMLOGRIAYVSGGSEEV	HKVN-----GNQILSSVRYG	QDLMSADDTQ 705
Db	714	PHYDAIDALLARIKTVAGQDMKV	TYMGVPREADKWSYNGILTSV	RYGTGANEATD-EG 772
Qy	706	TDLSTRSGLVTLVSDNPDLG--GSL	TVNMGRAHANQYRPLILGTG	DGVOSYLKDSDT 764
Db	773	TAETRTQGMVAVIASNNPNLKL	NEWDKLVNMGAAHKNQYRPLV	TTTKDGISRYLTDEEV 832
Qy	765	--NIVKYTDANGNLTFADDIKGY	STVDMSGYLAVWVPVGAQGDQ	QDVRVAADTNQKADGK 822
Db	833	PQSLWKKTDTANGILTFDMDIAG	SYNVQVSGYLAVWVPVGAQADQ	ADARTASKKKNASGQ 892
Qy	823	SLKTSAAALDSOVIYEGFSNFQ	DFANNADYTNKKIAENADFF	KKLGITISFEMAPQVVSAT 882
Db	893	VYESSAALDSOVIYEGFSNFQ	DFATRDDQYTNKVIKVNVL	FKEWGVTSEFELPPQVVSQ 952
Qy	883	DGSFLDSIIQNGYAFSDRYDLAMS	KNNKYKSGDDLANALKALHANG	IAQIAADVPPQIYO 942
Db	953	DGTFLLSIIQNGYAFSDRYDLAMS	KNNKYKSLKDLNLALRALHS	VNIQIAADVPPQIYN 1012
Qy	943	LPGEVVVTAKRTNSYGNPTFDAY	INNALYATNTWKSSGSDYQAQ	YGAGAFDELKAKYPDMF 1002

Db	1013	LPGEVVVTAKRTNSYGNPTFDAY	INNALYATNTWKSSGSDYQAQ	YGAGAFDELKAKYPDMF 1072	
Qy	1003	TVMNISTGKPIDPSTKIKOWEAKY	PNGTNVLGKAGYVLSDDATGKY	FVTNNGDFLPAS 1062	
Db	1073	ERVQISNGQKMTTDEKITKWSAKY	PNGTNILGRGAYVVLKOWASND	YLT-NRNGEIVLPK 1131	
Qy	1063	FTGDNQAKTFYDGTGMAYYSTSG	NKAVNSFIY-EGGHYYFFDKG	HWVTSYKAEDGN 1121	
Db	1132	QLVKNKSYTFGVSDANGTKFYST	SGYQAKNSFIQDENGWNWYF	DKRGYLVTAHEI--DGK 1190	
Qy	1122	DYVFLPNGIQERDALYQDAQNS	YYGRG--ILYK-----GDNWY	PFVDPNNAKTVPR 1174	
Db	1191	HVFLKNGIQLKDSIREBENG	NOYYDQTAQVNLNRYTTDQ	QNW-----R 1236	
Qy	1175	YDANNVMAIGYNNMYGQTYYP	DENGFOAKGOLLTDDKG--	THYFDEDDNGAMAKNFVNVG 1233	
Db	1237	YFDAGVMARGLVKIGDGOQ	PFDENGYQVKGKIVSAKDG	KLRYFDKDSGNAVNRPAQ--G 1295	
Qy	1234	D---DWYMDGNGNAVKGQY	PVNNQILYFNPETGTVQVKG	QFITDAQRTSYVDANSALK 1290	
Db	1296	DNPSWYFYGVBEFAKLTLG	KIGQIGQOQTLYFD--QDGKQV	KGKIVTLSDKSIRYFDANSGEWA 1354	
Qy	1291	SSGFPTPGSDWYVYAENG	VYVYKGFQVAENQDQWY	YFDQTTGKQAKGAHVGRDLYFNP 1350	
Db	1355	VCKF-----AEGA-----	-----KNEWYFDK--	TOKAVTGLQIKGQOQTLYFDQ 1391	
Qy	1351	DSGVQVKGDFATDES	SGTFSFYHNGDKVVG	FTTGNNAWYADNNGNLVKGFQ	EQEIDGK 1410
Db	1392	D-GKQVKGKVVTLADKS	IRYFPDADSGEMAVGFAE	GAKNEWYFDTGKAVTGLQ	ID-- 1448
Qy	1411	WTHFBEVTGQAKGAALV	NGQQLYFDVDSGTVQVKG	DFVTDGQNTSYVDVNSGDKK	VNGF 1470
Db	1449	-----KOTLYFDQD--	GKQVKGKIVTLSDKS	IRYFDANSGEWATNKF 1488	
Qy	1471	FTTGNNAWYADQCNLA	KGRKSIDNQDLYFDPATG	KQVKGOLVSDGRNYYFDSG	SGNN 1530
Db	1489	VEGSQNEWYFDAQKAV	TGLQVQOQTLYF--	TQDGKQVKGKVVNDVNGV	SRIFDANSMD 1547
Qy	1531	AKNRFVRIGD-QWYIF	GNDAATN 1553		
Db	1548	ARSKWQLEDGSMWY	FDRDGRQN 1571		

RESULT 8					
GTFFD STRMU					
ID	GTFFD STRMU	STANDARD;	PRT;	1462 AA.	
AC	P49331; O69383; O69386; O69389; O69392; O69398;				
DT	01-FEB-1996 (Rel. 33, Created)				
DT	28-FEB-2003 (Rel. 41, Last sequence update)				
DT	10-MAY-2005 (Rel. 47, Last annotation update)				
DE	Glucosyltransferase-S precursor (EC 2.4.1.5) (GTF-S) (dextranucrase)				
DE	(Sucrose 6-glucosyltransferase).				
GN	Name=gTFD; OrderedLocusNames=SMU.910;				
OS	Streptococcus mutans.				
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;				
OC	Streptococcus.				
OX	NCBI_TaxID=1309;				
RN	[1]				
RP	NUCLEOTIDE SEQUENCE.				
RC	STRAIN=GS-5;				
RX	MEDLINE=91100958; PubMed=2148600;				
RA	Honda O., Kato C., Kuramitsu H.K.;				
RT	"Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the glucosyltransferase-S enzyme."				
RL	J. Gen. Microbiol. 136:2099-2105 (1990).				
RN	[2]				
RP	NUCLEOTIDE SEQUENCE.				
RC	STRAIN=MT4239 / Serotype c, MT4245 / Serotype e, MT4251 / Serotype f,				
RX	MT4467 / Serotype e, and MT9148 / Serotype c;				
RA	MEDLINE=98231643; PubMed=9570124;				
RA	Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,				
RA	Kimura S., Hamada S.;				


```
QY 345 AAOQKWNMSTE--NFKNGDHLGGALLYNSDLTPWANSYRLNRTPTQODGTYKCYPT 401
Db 352 KTQPNWNSQTESDTSAGEKXHLGGALLYSNDSKTAAYANSYRLNRTPTSGPKPYE 411
QY 402 EGEGGYEFLSNDVNSNPVVAEQNLQHLNWNWDIVMGDKANFGDVRVADVNVN 461
Db 412 DNSSGGYDFLLANDINDNSNPVVAEQNLNHLNMYGSIIVANDPEANFGDVRVADVNVN 471
QY 462 ADLLQVYSNYFKDNYKVTDSEANALAHISILEAWSLNDNQYNEDTNGTALSINDSRLTS 521
Db 472 ADLLQIASDYLKAHYGVDSKSEKNAIINHLISILEAWSNDPQYNKDTKGAQLPINDKLRSL 531
QY 522 LAVLTQPCQRIDLSN-----LISESVKNERANDTAYGDTIPTYSFVRANDSEVQ 571
Db 532 LVALTR-PLEK-DASKNKEIRSGLEFVITNSLN-NRSAEGKNSERMANVIFIRAHDSVQ 588
QY 572 TWIAKIVKEDITNSDGYFTLTDQKDAFKIYNEDMAKNKTYTHYNIIPAAVALLISNNE 631
Db 589 TWIAKIIKAIQNPKTDLGTLTDLDELQAPKIYNEDMRQAKKYTQSNIPTAALMSNKD 648
QY 632 SVPRVYGYDLYDDGQYMAKKSPYDAIATMLQGRYAVVSGQSEBEVHKVNGNQ----- 686
Db 649 SITRLYGYDMYSDGQYMATKSPYDAIDITLLKARIKYAAGQDMKITVVEGDKSHMDW 708
QY 687 ---ILSSVRYGQDLMSADDTQGTDLRSRTSLGLVTLVNSNDPNLDLG-GDSLTVNMGRAHNO 742
Db 709 YTGVLTSVRYGTCANSEATD-QGSEATKTQGMVITNSNPSLKNQNDKVINMGTAHKNQ 767
QY 743 AYRPLILGKDGQVQSYLKDSDR-NIVKYTDAMGNLTFTADDIKGYSTVDMSGYLAVWVPV 801
Db 768 EYRPLLLTLYKDLGTSYTSDAAKSLYRKTNDKGELVFDASDLOGYLNPQVSGYLAVWVPV 827
QY 802 GAKGDQVRVAADTNQADGKSLKTSAAALDSQVIYEGFSNFQDFANNDADYTNKKIAENA 861
Db 828 GASDNQDVRVAASNKANATQVYESSALDSQLIYEGFSNFQDFVTKDSYTNKKIAQNV 887
QY 862 DFEKLGITSPEMAPQYVSGATGSLDSIIQNGYAFSDRYDLAMSKNNKYGSKDDLANAL 921
Db 888 QLFKSGVTSFEMAPQYVSGESDGLDSIIQNGYAFSDRYDLAMSKNNKYGQQDMINAV 947
QY 922 KALHANGIOAIDWPDQIYQYLPGBERVVTAKTNSYGNPTFDAYINNALYATNTKSSGD 981
Db 948 KALHKSIGIOAIDWPDQIYQYLPGBERVVTAATRVNDYGEYRKDSEIKNTLYAANTKSNKGD 1007
QY 982 YQAQYGGAFDLBKAKYPMFTNMISTCKPIDPSTKIKQWBAKYPNGTNVLKGGAGYVL 1041
Db 1008 YQAKYGGAFSLSELAAYPSIFNRTQISNGKKIDPSEKITAWKAKYPNGTNILGRGVGYL 1067
QY 1042 SDDATGKYFTVNEGDFLPASPTGQDQNAKTGYDGTGMAYYSTSGNKAIVSFIYEG-CH 1100
Db 1068 KDNASDKYFELKGNQTYLPKQMT-NKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKN 1126
QY 1101 YYYFDDHGMVTSYKAEQDNDYVPLPNGIQMRDAIYQDAQNSYYVYGRGTGILYKGDWNY 1160
Db 1127 WYTFDNGHMVVG-LOHNGEYQVYFLSNGVQLRESFLENADSGKNYFGLHGRY-SNGY 1184
QY 1161 PFVDPNNAKTYFPRYDANNVMAIGYRNNYMGOTYTPDENGFOAKGOLLTDDXG-THYFDE 1219
Db 1185 SF-----DNDSKWRYPDASGVMAVGLKTINGNTQYFDQGYQVKGAWITGSDGKKRYFDD 1239
QY 1220 DNGMAKKNFVN-VGDDWYMDGNGNAVKGOYFVNNQIILYFNPETGVQVKGQFITDAQCR 1278
Db 1240 GSGNMAVNRFPANDKNGDWTYLSNDSGLALVGQVQTINGKTYFYG-QDGKQIKGKIITD-NGK 1297
QY 1279 TSYDANSALKSSGFFTPNGSDWYVAENGYVYKGFQVAENQDOWYVDTQTTGKQAKGA 1338
Db 1298 LKXFLANSSELARNIPAT-----DSQNNWYFYG-SDGVAVTGS 1334
QY 1339 AKVDGRDLYFNPDSGVQVKGDFATDSBGNSTSFYHGDNGDKVYGGFFTTGNNAWYADNNG 1398
Db 1335 QTIAGKKLYFASD-GKQVKGVSFT-YNGKVHYTHADSGELQVNRFEADKDGNNWYLDNSG 1392
QY 1399 NLVKGQFQIDGKWHYFDEVTGQAKGAALVNGQQLYFVDVDSIGIYQVKGDFVTDQGNSTSY 1458
```

```
Db 1393 EALTSQRI-----NGQRVFF-TREGKQVKGDVAYDERGLLRY 1430
QY 1459 DVNSGDKKXNGVFPTTGDNAWYADQGGLAKGRK 1492
Db 1431 DKNSGNMVKVVT-----LANGRR 1450

RESULT 9
Q59983 9STRE
ID Q59983 9STRE PRELIMINARY; PRT; 1590 AA.
AC Q59983
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-WAR-2004 (TREMBLrel. 26, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5).
GN Name=gtfI;
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OC NCBI_TaxID=1310;
RN (1)
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OM2176;
RX MEDLIN=94146405; PubMed=8312602;
RA Sato S., Inoue M., Handa N., Aizawa Y., Isobe Y., Katayama T.;
RT "DNA sequence of the Glucosyltransferase gene of serotype d
RT Streptococcus sobrinus.";
RL DNA Seq. 4:19-27(1993).
RN (2)
RP PROTEIN SEQUENCE.
RX MEDLINE=91224988; PubMed=1827439;
RA Mooser G., Hetta S.A., Paxton R.J., Shively J.E., Lee T.D.;
RT "Isolation and sequence of an active-site peptide containing a
RT catalytic aspartic acid from two Streptococcus sobrinus alpha-
RT glucosyltransferases.";
RL J. Biol. Chem. 266:8916-8922(1991).
DR EMBL; D13858; BAA02976.1; -; Genomic_DNA.
DR PIR; A39841; A39841.
DR HSP; P06653; IHCK.
DR GO; GO:0047849; F:dextranucrase activity; IEA.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . ; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Glycosyltransferase; Signal; Transferase.
FT SIGNAL 1 38 Potential.
FT CHAIN 39 1590 Glucosyltransferase-I.
FT SEQUENCE 1590 AA; 175956 MW; C3C83A57CF3C2B0E CRC64;

Query Match 42.5%; Score 3502.5; DB 2; Length 1590;
Best Local Similarity 45.4%; Pred. No. 2.1e-144;
Matches 749; Conservative 252; Mismatches 482; Indels 167; Gaps 40;

QY 1 MEKLLHYKLKHKGHWVTIAVASIG-LVSLVGAGTVSAEDKVANDTTAAQTVDGTGQDQ 59
Db 1 MEKNVRFKKHKKRQVWTLVSASATMLASALCASVASADTDTASDDSNQAVV---TG-DQ 56
QY 60 ATTND--NTNTTDTTADQSNQDQSGSDNNQDAKQDNTANTDRNQADNSOTDNN 117
Db 57 TTNNQATQTSIAATATSEQSASTD---AATDQASAAEQTQGTASTD--TAAQTFTTAN 111
QY 118 QATDQATSPATDG-TSVQRDAAVATAADQSGQTAPSEQKSAALSILNKLIDGKYKY 176
Db 112 EAKVPTENEGQFTDEMLAEAKNVATA---ESDISPDLAK-----MSNVKQVDGKYKY 163
QY 177 VOADGSYKQNFAYTNGQMLYFDSITGALSSSTSTYSFSGTNTNLVDD---FSSHNKAYDS 233
Db 164 YQDQGNVKNKFAVSGDKIYYFD-ETGAYKDTSKVDADKSSSAVSQNAIFAANNRAYST 222
QY 234 TAKSFELVNGYLTANSWYRPAGILRNGQTWEASNENDLRPVLMSWMPDKDQTVVYVYN 293
```

[illegible]

Db 1380 DKAGNWFYLGK-----DGAANVTGAOTIRGQKLYFKA-NGQQV 1311
 QY 1357 KGDPATDESNGNTSYHGNDGKGVGGVFFTTGNNAAWYADNNG-----NLVKG-----1403
 Db 1316 KGDIVKGTGDKIRYDAKSGEQVFNKTVKAADGKTYVIGNDGVAVDPVSVYKGTQFXDASG 1375
 QY 1404 ---FQETDGGK-----WY-----HFDEV---TGQQAAGKALVNGQQLYFDVDSGIGQVKGDFV 1448
 Db 1376 ALRFYNLKGQLVLTGSGWYETANHDWYVIOGSKALTGEQTINGQHLFKED-GHQVKGQLV 1434
 QY 1449 TDGCGNTSYDVSNGDKKVGNGFFTTGDNAWY-----1480
 Db 1435 TGTDGKRYDYDANSGDQAFNKSVTNGKTYFNGNDGTAQTAGNPKGQTQKDGSDIRPYSM 1494
 QY 1481 -----ADGQ-----GNLAKGRKSIDNQDLYFDPATGKQVKGQLV-SIDGRN 1520
 Db 1495 EGQLVTGSGWYENAQQGLVYKNGKVLTLGQTVGSQRVYFD-ENGIOAKGKAVRTSDGKI 1553
 QY 1521 YYFDSGSGNAKRFVIRIGDQWLYFNGDGA 1550
 Db 1554 RYFDNSGSGMITNQWKFVYQGYFNGDGA 1583

 RESULT 10
 GTFB_STRMU
 ID AC P08987; O69381; STANDARD; PRT; 1476 AA.
 AC P08987; O69381; O69387; O69390; O69396;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-MAY-2005 (Rel. 47, Last annotation update)
 DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
 DE (Sucrose 6-glucosyltransferase)
 GN Name=gtfB; OrderedLocusNames=SMU.1004;
 OS Streptococcus mutans.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1309;
 RN [1]
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=GS-5;
 RX MEDLINE=87308013; PubMed=3040685;
 RA Shiroza T., Ueda S., Kuramitsu H.K.;
 RL "Sequence analysis of the gtfB gene from Streptococcus mutans.";
 RT J. Bacteriol. 169:4263-4270(1987).
 RN [2]
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=MT4239 / Serotype c, MT4245 / Serotype e, MT4251 / Serotype f,
 RC MT4467 / Serotype e, and MT8148 / Serotype c;
 RX MEDLINE=98231643; PubMed=95701124;
 RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
 RA Kimura S., Tamada S.;
 RT "Molecular analyses of glucosyltransferase genes among strains of
 RT Streptococcus mutans.";
 RL FEMS Microbiol. Lett. 161:331-336(1998).
 RN [3]
 RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RC STRAIN=UA159 / ATCC 700610 / Serotype c;
 RX MEDLINE=22259063; PubMed=12397186; DOI=10.1073/pnas.172501299;
 RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
 RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
 RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,
 RA Ferretti J.J.;
 RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
 RT pathogen.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
 CC -1- FUNCTION: Production of extracellular glucans, that are thought to
 CC play a key role in the development of the dental plaque because of
 CC their ability to adhere to smooth surfaces and mediate the
 CC aggregation of bacterial cells and food debris.
 CC -1- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-
 CC fructose + (1,6-alpha-D-glucosyl) (n+1).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha


```
QY 765 NIVKYTDANGNLFTTADDIKGYSTVDMGSLAVWVPVGVKADQDVRVAADTNQKAGKSL 824
Db 760 GLVRYTNDRGELFTTADIKGYANPQVSGYLGVAADQDVRVAASTPSTDKGSV 819
QY 825 KTSAAALDSOVIYEGFNFODFANNDADYTNKKIAENADPFKLGITSEFEMAPQYVSATDG 884
Db 820 HQNALDSRVWFEGFNFQAFATKBEYTNVJAKNVDKFAEAGVTDFAEPQYVSSTDG 879
QY 885 SFILDSIQNGYAFSDRYDLAMSKNNKYGSKDDLANALKALHANGIOAIAADWPDQIYQLP 944
Db 880 SFILDSIQNGYAFSDRYDLAMSKNNKYGSKDDLANALKALHANGIOAIAADWPDQIYQLP 939
QY 945 GBEVWTAKTNSVGNFTFDAYINNALYATNTKSSGSDYQAYCGAFDELKAKYPMFTV 1004
Db 940 EKEVWTAKTNSVGNFTFDAYINNALYATNTKSSGSDYQAYCGAFDELKAKYPMFTV 999
QY 1005 NMISTCKPDPSPKIKQWAKYFNGTNVLGKAGYVLSDDATGKYFTVNEGDG--FLPAS 1062
Db 1000 KQISTGVNPDSPVKIKQWAKYFNGTNVLGKAGYVLSDDATGKYFTVNEGDG--FLPAS 1059
QY 1063 FTGDQNAKTGFYDGTGMAYSTSGNKAVNSFIYEGGHYYFDKOGHMYTGSYKABDGDND 1122
Db 1060 LL-NODSQVGSYDGGKGYVYSTSGYQAKNTFISEGDKWYFDNNGYMYTGA-QSINGYN 1117
QY 1123 YYPELNGIOWRAIDYQDAQNSYYIGRTGLYKGDWNYFPVDPNNAKNTVRYFDANNVM 1182
Db 1118 YYPELNGIOWRAIDYQDAQNSYYIGRTGLYKGDWNYFPVDPNNAKNTVRYFDANNVM 1169
QY 1183 AIGYRMYGTYVYFDENGFOAGKOLLTDDKG--THYFEDENGAMAKNFV-NVGDDWYND 1240
Db 1170 SVGLTWIDQVQVYFDENGFOAGKOLLTDDKG--THYFEDENGAMAKNFV-NVGDDWYND 1229
QY 1241 GNGNAVKGQPVNNQILYFNPETGVQVKGQFITDAQRTSYDANSALKSSGFFTPNGS 1300
Db 1230 EDGAATVGSQTNGHLYFR-ANGVQVKGFEVTDYGRISYDANSNGDQIRNRF----- 1282
QY 1301 DWYBAENGVIYKGFQVAENQOQYVYFDOTTKQAKGAAGVQDGRDLYFNPDSGVQVKGDF 1360
Db 1283 -----VRNAOQGWYFED--NNGYAVTGARTINGQHLYFRA-NGVQVKGFE 1324
QY 1361 ATDESGNTSFYHGDNGDKVVGFFFTGNNAWYADNNGNLVKGFEIDGKWHFDEVTQ 1420
Db 1325 VTDRHGRISYDNGSGDQIRNRFVRNAOQGWYFONNGYAV----- 1365
QY 1421 QAKGAALVNGQQLYFVDSGQVKGDFVTDGQNTSYDVSNGDKKVGFFTTGDNWY 1480
Db 1366 --TGARTINGQHLYFRA-NGVQVKGFEVTDYGRISYDANSNGDQIRNRFVRNAOQGWY 1422
QY 1481 ADGQGNLAKGRKSIDNQDLYFPDPAKQVKGQVLSID-GRNYYFDSGSGNMAK 1532
Db 1423 FDNNGYAVTGARTINGQHLYF-RANGVQVKGFEVTDYGRISYDANSGERVR 1474

RESULT 11
Q55263_9STRE PRELIMINARY; PRT; 1590 AA.
AC Q55263;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-JAN-1998 (TremBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)
DE GTF-I.
GN Name=Glucosyltransferase;
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1310;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 33478;
RA Sato S.;
RT "DNA and amino-acid sequences of water-insoluble-glucan synthetase
RT produced from Streptococcus sobrinus ATCC 33478."
RL Kagoshima Daigaku Shigakubu Kiyo 16:23-29(1996).
```

```
RN [2]
RP PROTEIN SEQUENCE.
RX MEDLINE=91224988; PubMed=1827439;
RA Mooser G., Hefta S.A., Paxton R.J., Shively J.E., Lee T.D.;
RT "Isolation and sequence of an active-site peptide containing a
RT catalytic aspartic acid from two Streptococcus sobrinus alpha-
RT glucosyltransferases."
RL J. Biol. Chem. 266:8916-8922(1991).
DR EMBL; D63570; BAA09792.1; -, Genomic_DNA.
DR PIR; A39841; A39841.
DR HSSP; P06653; 1GYM.
DR GO; GO:0003250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 3.
DR Pfam; PF02324; Glyco_hydro_70; 1.
SQ SEQUENCE 1590 AA; 176058 MW; 9DF7A3F2C6B4FD43 CRC64;

Query Match 42.3%; Score 3481.5; DB 2; Length 1590;
Best Local Similarity 45.3%; Pred. No. 1.8e-143;
Matches 748; Conservative 249; Mismatches 487; Indels 167; Gaps 41;
```

```
QY 1 MEKKLHYKLHKYKHWVTIAVASIG-LVSLVGAGTVSAEDKVANDTTAAQATVGVDPGQDQ 59
Db 1 MEKNVRFKMKHKRWVTLVSASATMLASALCASVASADTDITASDDSNQAVV---TG-DQ 56
QY 60 ATTNDAA--NTNTTDTADQSANNTQDOAGSQNNQDQAKQDQTANTDRNQADNSOTDNN 117
Db 57 TTNNQATQTSIAATATSEQASSTD---AATQASAAEQTQGTASTD--TAAQTITNAN 111
QY 118 QATQATSPATDG-TSVQRDDAANVATAADQSGQAPSEKSAALSNDLVKIDGKYV 176
Db 112 EAKWVPTENENQGTDELAENKAVATA---ESDFFPSDLAK-----MSNVKQVDGKYV 163
QY 177 VQADSGYKKNFPAITVNGQMLYFDSGTGALSSTSTYSFSGQITNLVDD---FESHNKAYDS 233
Db 164 YDQDGNVKNKFAVSGDKIYFDF-ETGAYKDTSKVDADKSSVASQNAITFAANNRAYST 222
QY 234 TAKSPELVNGYLTANSWYRPAGILRNGOTWEASNENDLRPLVMSWMPDKDQTVAYVYVYN 293
Db 223 SAENFEADVNLITADSWYRPKSLKDGKTWESGKDDFRPLLMANWPDTETRKYVYVYN 282
QY 294 KYLSANETEVTNETSQVDLNKEAQSIQTKIEQKITSDNSTOWLTAMEAFVAAQPKWNMS 353
Db 283 LVVGIDKT-YTAETSQADLTAAAEVLQARIEQKITTEQNTKWLREASIAFAVTKPQWNGE 341
QY 354 TENFKGDLHQQGALLYTN-SDLTPWANSDFYLLNRTPTQDGT-KKYPTGEGE---GGY 408
Db 342 SEK-PYDDHLQNGALKFDNQSDLTPTQTSNYRLLNRTPTNQTGSLDSRFTYNANDPLGGY 400
QY 409 EFLLSNDVDNSNPVQAEQLNQLHVMWGDIVMGDKDANFDGVVDAVDNVNADILQVY 468
Db 401 EFLLANDVDNSNPVQAEQLNQLHVMWGDIVMGDKDANFDGVVDAVDNVNADILQVY 460
QY 469 SNYFKDNYKVTDSEANALAHISILEAWSLNDQNYNEDTNGTALSIDNSRLTSLAVLTQ 528
Db 461 SDYLKAAVGDIDKNNKANNHVSIVEASDNDTPYLHDDGDNLMNMDNFKRLSMLSLAKP 520
QY 529 PQQRIDLNLSISESVNKERANDTAYGDTIPYVSFVRAHDSSEVQTVIAKIVKSKIDTNSDG 588
Db 521 LDKRSGLNPLIHNSL-VDREYDDREVEVTPSYSFARAHDSVEQDIIRDIKAEINPNSFG 579
QY 589 YTFLLDQLKDAFKIYNEDMAKNKTYTHYNTPAAVALLSNMESVPRVYVGYDLYDDGGY 648
Db 580 YSFQEEEDQAFKIYNEDLKKTKYTHYNVPLSYTLTLTKNGSIPRVYIGMDFDDGGY 639
QY 649 MAKSPYYDAIATMLQGRYATVSGQSEVHKVNGNNOILSSVRYGQDLMSADDTQGTDL 708
Db 640 MANKTVNYDAIESLLKARMKYVSGGQAMQNYQI-NGEILTSVRYGKALKQSD-KGDAT 697
QY 709 SRTSGLVTLVNSNDPNLDLGGDSLTVMNGRAHANQAYRPLILGTQKGVOSYKSDSTN--- 765
Db 698 TRTSGVGMGNQPNFSLDGKVALNMGAAHANQAYRALMVSTKQGVATYATDADASKAG 757
```

```
QY 766 IVKYYTDANGNLFTADDIKGYSTVDMSGVLAWVPVGAQGDQDVRVAADTNQADCKSLK 825
DB 758 LVKRTDENGVLPLNDLKGVANPQVSGFLQWVPVGAADDQDIRVAASDTASTDGKSLH 817
QY 826 TSAALDSQVIYEGFSNFQDFANNADYTNKKIAENADFFKKLGITSFEMAPQVVSATDGS 885
DB 818 QDAMDSDRVMPGFSNFQSFATKEEYTNVANNVDKFKVSGWITDFEMAPQVVSSTDOQ 877
QY 886 FLDSIIQNGYAFSDRYDLAMSKNNKYGSKDDLANALKALHANGIQAIADVPDQYQLPG 945
DB 878 FLDSVIQNGYAFSDRYDLGMSKANKYGTADQLVKAIKALHAKGLKVMADVPDQMYTFPK 937
QY 946 EEWVTAKRNSVGNPTFDAYINNALLYATNTKSSGSDYQAOYGAGFLDELKAKYPMFTVN 1005
DB 938 QEVVTVTRDKFPGTAGSQINHSLSYVTDTKSGDDYQAKYGAFLDELKERYPELFTKK 997
QY 1006 MISTGKPIDPSTKIKOWEAKYFNGTVLKGAGYVLSDDATGKYFTVNGSGDFPLPASFTG 1065
DB 998 QISTGQAIQDPSVKIKQWSAKYFNGSNILRGADYVLSQVSNKYFNVASDTLFLPSSLLG 1057
QY 1066 DQNAKTGFYDGTGMAYYST-SGNKAVNSFIYEGGHYYFYDGDGHMVTGSYKAEDGNDYY 1124
DB 1058 -KVVESGIRYDGGKGIYNSSTGTDQVKASFIIEAGNLYYFGKGYVMTGA-QTINGANYF 1115
QY 1125 FLPLNGIQMDAIYQDAQNSYYGRTGILYK-----GDNWYFVDPNNANKTVRYFPA 1178
DB 1116 FLENGTALRNTIYTDAGNSHYANDGKRYENGYQQFGNDW-----RYFKD 1161
QY 1179 NNMAIGYRNMVGTOTYFDENGFOAKGQLL-TDDKGFHYFDEBNGAMAKNFV-NVGDDW 1236
DB 1162 GN-MAVGLTVDGNVQYFKDQGVQAKDKIIVTRDGKVRFPDQNGNAVNTFIADKTGH 1220
QY 1237 YYMDGNGNAVQGYPVNNQILYFNPETGYQVKGQFITDQGRTSYYDANSALKSGGFT 1296
DB 1221 YYLGKDGVAVTGAQTVGKQKLYPE-ANGQVKGDFVTSHEGKLYFYDVSQDGMWTDFT 1279
QY 1297 PNGSDWYYAENGYYVYKGFQVAENQDQWYFDTTQKQAKAAKVDGRDLYNPNDSGVQV 1356
DB 1280 DKANGNYFLGK-----DGAAYSGAGTIRGQKLYFKA-YGQOV 1315
QY 1357 KGDFAFDGSGNTSFYHGDNDGKVVGGFFTTGNNAWYADNNG-----NLVKG----- 1403
DB 1316 KGDIVKGTOKIRYDAKSGEQVFNKTVKAADGKTVIGNGVAVDPSVVGKQTFKDSG 1375
QY 1404 ---FQIBDGK-----WY-----HFDEV---TGQOAKGAALVNGQQLYFDVDSGQVKGDFV 1448
DB 1376 ALRFYNLKQLVGTSGWYETANHDWYVYIOSGRKALTGEQTNGHLYFKED-GHQVKGQLV 1434
QY 1449 TDQGGNTSYVDVNSGDKKNGFFTTGDNWY----- 1480
DB 1435 TRTDGKVRYYDANSQDQAFNKSVTYVNGKTYFYFNGDGTAGTGNPKGQIFKDGSLVRYSM 1494
QY 1481 -----ADQO-----GNLAKGRKSTDNQDLYFDPATGKQVKGQLV-STDGRN 1520
DB 1495 EGQVLVIGSGWYSNAQQWLYVKNKVLTLGLQTVGSORVYFD-ENGIOAKGKAVRTSDGKI 1553
QY 1521 YYPDSGSGNMAKNRFRVIRIGDQWYIFGNDGAA 1551
DB 1554 RYFDENSGSMITNQWKFVYQYQYFYFNGDGA 1584

RESULT 12
ID GTF1_STRDO STANDARD; PRT; 1597 AA.
AC P11001;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-1) (Dextranucrase)
DE (sucrose 6-glucosyltransferase).
GN Name=gtf1;
OS Streptococcus downei (Streptococcus sobrinus).
```

```
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1317;
RN [1]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC STRAIN=MFE28;
RC MEDLINE=87308014; PubMed=3040686;
RX Ferretti J.J., Gilpin M.L., Russell R.R.B.;
RA "Nucleotide sequence of a glucosyltransferase gene from Streptococcus
RT sobrinus MFE28.";
RL J. Bacteriol. 169:4271-4278(1987).
CC -!- FUNCTION: Production of extracellular glucans, that are thought to
CC play a key role in the development of the dental plaque because of
CC their ability to adhere to smooth surfaces and mediate the
CC aggregation of bacterial cells and food debris.
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-
CC fructose + (1,6-alpha-D-glucosyl) (n+1).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
CC water-soluble glucans and some 1,6-linkages), GTF-S synthesizes both
CC forms of glucans.
CC -!- SIMILARITY: Belongs to the glucosyl hydrolase 70 family.
CC -!- SIMILARITY: Contains 19 cell wall binding repeats.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL: M17391; AAC63063.1; -; Genomic DNA.
DR InterPro: IPR002479; Cell_wall_bd_put.
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; CW_binding_1; 4.
DR Pfam: PF02324; Glyco_hydro_70; 1.
KW Dental caries; Glycosyltransferase; Repeat; Signal; Transferase.
FT SIGNAL 1 38 Potential.
FT CHAIN 39 1597 Glucosyltransferase-I.
FT REPEAT 1099 1132 A repeat.
FT REPEAT 1163 1213 AC repeat.
FT REPEAT 1227 1277 AC repeat.
FT REPEAT 1292 1342 AC repeat.
FT REPEAT 1352 1399 B repeat.
FT REPEAT 1406 1455 AC repeat.
FT REPEAT 1465 1512 B repeat.
FT REPEAT 1582 1597 A repeat (incomplete).
FT REGION 39 1050 Catalytic (approximate).
FT REGION 1099 1597 1,25 A, 2 B and 5 AC repeats.
FT REGION 1099 1597 Glucan-binding (approximate).
FT SEQUENCE 1597 AA; 177080 MW; B9E86A200868798B CRC64;

Query Match 42.3%; Score 3481; DB 1; Length 1597;
Best Local Similarity 44.6%; Pred. No. 1,9e-143;
Matches 739; Conservative 261; Mismatches 484; Indels 174; Gaps 40;

QY 1 MEKKLHYKLHKVKKHWVTIIVASIG-LVSLVGAGTVSAEDKVANDTTAQTATVGVDTGDDQ 59
DB 1 MEKNERFKMKVKKRWVTISVASATMLASALCASASADTETVSDSNQAVLTA-----DQ 56
QY 60 ATTND--ANTNTTDTDTADQSANTN--ODQA--GSDSNNDQQAQKQDTANTDRNQADNSOT 114
DB 57 TTTNTDTEQTSVAATATSEQSASTDAATDQASAAEQTQGTASTD--TAAQTFTT 114
QY 115 DNNQA-----TDQATSPATDGTQVQRDAANVAATAADQEQATSPSEQEKAALSLDNVKLI 170
DB 115 NANEAKWVPTENENQVFTD---EMLAEXKNVATA---ESNSIPSLAK-----MSNVKQV 163
QY 171 DGKYYYQADGSYKKNFAITVNGQMLYFDSDFGALSSSTYSFSQGTNNLVDD---FSSH 227
DB 164 DGKYYYQDQGNVKNKFAVSGEKIYFDD-ETGAYKDTSKVEADKSGSDISKEETTTFAAN 222
```



```
QY 228 NKAYDSTAKSFELVYLTANSYVRPAGILRNQGTWEASNENDLRPLVMSWPKDQTQVA 287
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
223 NRAYSTSAENFEADINYLTAADSYRPRKSIILDKGTWTESSKDDFRPLLAAMWPDTEKRN 282
QY :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
288 YVNMKYLISANETEVNTSOTVDLANKEAQSIOTKLEQKITSNSTOWLRTAMEAFVAAQ 347
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
283 YVNMKLVGIDKT-YTAETSOADLTAALAEVQARIEQKITTEQNTKWLUREALISAFVKQ 341
QY :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
348 PKWNMSTENFNKGDLHGGGALLYTN-SDLTPWANSDYRLINRPTTQDDGT-KKYFTEGGE 405
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
342 POWNGESEK-PYDDHLONGALKEFDNQSDLTPTQSNYRLINRPTTQDGLSDSRFTYNAN 400
QY :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
406 ---GGEFFLLSDVNSNPVQAEQLNQLHYLMWGDIVMGDKXANPDFGVVDVAVNVA 462
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
401 DPLGGYELLANDVNSNPVQAEQLNQLHYLMWGDIVMGDKXANPDFGVVDVAVNVA 460
QY :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
463 DILQVTSNYFKDNYKYVDSEANALAHISILEAWSLNDNOYEDTNGTALSIDSNSRLTSL 522
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
461 DILQISSDYLKAYGIDKKNKANNHVSIVEAWSNDTPTYLHDDGONLMNMONKFLSML 520
QY :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
523 AVLTKOPGQRIDLSNLISVSKERANDTAYGDTIPTYSFVRAHDSEVQTVIAIKVEKI 582
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
521 WSLAKPLDKKSLNPLIHNSL-VREVDDREVETVPSYFARAHDSVQDLIRDIKAEI 579
QY :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
583 DTNSDGYTTLQDLQDAFKIYNEDMAKVNKTYTHYNI PAAYALLJNSMESVPRVYGDLY 642
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
580 NPNFAGYSPTQDEIDQAFKIYNEDLKTKTKYTHYVPLSYTLTLLTKGSI PRVYGDWF 639
QY :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
643 TDDGQVMAKSPYDAIATMLQRIAYVSGQSEEVHKYNGNNOILSSVRYQDLSADD 702
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
640 TDDGQVMAKTVNYDAIBELLKARKMYVAGGAMQNYQI-NGEILTSVRYKGALQKSD 698
QY :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
703 TQGTDLSTRSGLVTLSVNSDNLGLGDSLTVMNGRAHANQAYRPLILGTGQVQSYLKDS 762
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
699 -KGDATRTSGVGVWNGQPNFSLGKVVALNNGAAHANEYRALMVSTKDGVAITYATDA 757
QY :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
763 DTN----IVKYTPANGNLTFTADDIKGYSTVDMSGYLAVWVPVGAQDQDVRVAADTNOKA 819
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
758 DASKAGLVKRTDENGLYFLNDDLKGVANPQVSGFLQVWVPVGAADQDQIRVAASDTAST 817
QY :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
820 DGSLKTSALDSOVIYEGFSNPQFANNADYTNKKIAENADFPKGLGITSFEMAPQV 879
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
818 DGKSLHQDAMDSRVWFEGFSNPFQSFATKEEBYTNVVIANNVDFKFSWGITDFEMAPQV 877
QY :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
880 SATDGSFLDSIIQNGVAFSDRYDLAMSNNKYGSKDOLANALKALHANGIOAIADWVPDQ 939
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
878 SSTDQQLFDSVLQNGVAFDTRDVLGSKANKYGTADQLVKALKALHAKGLKVMADWVPDQ 937
QY :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
940 IYQLPGEVVTAKRTNSYGNPTFDAYINNALYATNTKSGSDYQAYGGAFIDELKAKYP 999
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
938 MYTFPKQEVWTVTRTDKFKPIAGSOINHSLVYTDTKSGDDYQAKYGGAFIDELKAKYP 997
QY :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1000 DMFTVNMISTGKPIDPSIKIOWEAKYFNCTVILKGAGYVLSDDATGYFTVNGDPL 1059
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
998 ELFTKKQJSTQGAIDPSVKIKOMSAKYFNGLRGADYVLSQASNNKLYNSVDKDLFL 1057
QY :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1060 PASFTGDQNAKTFYVYDGTGMAY-YSTSGNKAVNSFIYEGGHYVFDKGHMTVGSYKAE 1118
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1058 PKYLLG-QVVEGIRFDGTGYVNSSTTGEKVTDSFITEAGNLYTFGQDGYMVTQAQNIK 1116
QY :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1119 DGNDFYFLPNGIQMRDAIYODAQNSYYGYRTGILYK-----GDNMYVFDVPPNANKT 1171
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1117 -GSNYFFLANGAALNTVYTDAGQGHYVNGDKRYENGYQFGNDSW----- 1163
QY :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1172 VFRYFDANNVMAIGVRNMYGQTYTFDENGFOAKGQLL-TDDKGTHYFDEEDNGAMAKNFV 1230
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1164 --RYP-KNGVMALGLTVDGHVQVYFDKGVQAKDKIIIVTRDGKRYVFDQHNNGNAVNTFV 1220
QY :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1231 -NVGDDWYMDGNGNAVKGQVFNNOILYENPETGVQVKGQFITDAQGRTSYDANSAL 1289
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1221 ADKTHGYLYLKGQGVAVVTAQTVGQHLHYFE-ANGQVQKGVDFVTAQDKGLFYFVDVSDGM 1279
QY :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1290 KSSGFTFPNGSDWYVYVYKPKQVAENQDQWYFDTTGTGQAKGAAKVGDGRDLYFN 1349
```

Db 1280 WINTFTIEDKAGNWFYLGK-----DGAAVTGAQTIKGQKLYPK 1316

QY 1350 PDSGVQVKGDFATDESNTSFYHGDNGDKVVGFFTTGNNAWYADN----- 1396

Db 1317 A-NGQQVKGDI VKDAGKIRIYDQAOTGEQVFNKSVNGKTYFSGDGTAAQOANPKGT 1375

QY 1397 --NGNLVKGFOEIDGK-----WYHFDE-----VTGQQAAGALVNGQOLYFDDVSGIQ 1442

Db 1376 FXDGGVLAFLYNLEGOYVSGSGWYETAHEHWYVYKSKVLTTGAQTIGNQRYVYK-DNGHQ 1434

QY 1443 VKGDFVTDQCGNTSYDVNSGDKKVNKGFPTTGDNAWY----- 1480

Db 1435 VRGQLVTGNDGKRLRYDANSQGDAFNKSVTVANGKTYFSGDGTAAQOANPKGTQFXDGG 1494

QY 1481 -----ADGO-----GNLAKGRKSIDNODLYFDDPATGKQVKGQLY 1514

Db 1495 VLRFYNLEGOYVSGSGWYKNAQGMVLYKDGKVLTLGLTVGNQKVYFD-KNGIOAKGKAV 1553

QY 1515 -SIDGRNYFVDSGSGNMKNKRVFRIGDQWIFYGNDGAA 1551

Db 1554 RTSDDKVRVFDENSGSGSMITNQMKFYVGYQYFSGDGA 1591

RESULT 13

GTF2_STRDO

ID_GTF2_STRDO STANDARD; PRT; 1592 AA.

AC P27470;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 13-SEP-2005 (Rel. 48, Last annotation update)

DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranase)

DE (Sucrose 6-glucosyltransferase).

OS Streptococcus downei (Streptococcus).

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI_TaxID=1317;

RN [1]

RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].

RC STRAIN=6715 / Serotype G;

RX MEDLINE=91123227; PubMed=1704006;

RA Abo H., Matsumura T., Kodama T., Ohta H., Fukui K., Kato K., Kagawa H.;

RA "Peptide sequences for sucrose splitting and glucan binding within Streptococcus sobrinus glucosyltransferase (water-insoluble glucan synthetase)".

RT J. Bacteriol. 173:989-996(1991).

CC -!- FUNCTION: Production of extracellular glucans, that are thought to play a key role in the development of the dental plaque because of their ability to adhere to smooth surfaces and mediate the aggregation of bacterial cells and food debris.

CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-fructose + (1,6-alpha-D-glucosyl) (n+1).

CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both forms of glucans.

CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.

CC -!- SIMILARITY: Contains 16 cell wall binding repeats.

CC This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.

CC EMBL; D90213; BAAL4241.1; -; Genomic_DNA.

DR HSSP; P06553; IGW.

DR InterPro; IPR002479; Cell_wall_bd_ptc.

DR InterPro; IPR003318; Glyco_hydro_70.

DR Pfam; PF01473; CW_binding_1; 3.

DR Pfam; PF02324; Glyco_hydro_70; 1.

KW Dental caries: Glycosyltransferase; Repeat; Signal; Transferase.	
FT SIGNAL	1 38 Potential.
FT CHAIN	39 1592
FT REPEAT	1093 1142 Glucosyltransferase-I.
FT REPEAT	1158 1207 1.
FT REPEAT	1222 1272 2.
FT REPEAT	1287 1337 3.
FT REPEAT	1402 1451 4.
FT REPEAT	1514 1563 5.
FT REPEAT	1577 1592 6.
FT REGION	39 1044 7 (incomplete).
FT REGION	1093 1592 Catalytic (approximate).
FT REGION	1093 1592 7 x tandem repeats.
FT REGION	1093 1592 Glucan-binding (approximate).
SQ SEQUENCE	1592 AA; 176168 MW; BC0A66D079351ECF CRC64;
Query Match	
Best Local Similarity 42.2%; Score 3472.5; DB 1; Length 1592;	
Matches 741; Conservative 254; Mismatches 496; Indels 153; Gaps 39;	
QY 1	MEKKLHYKLVKVKHWTIAVASIG-LVSLVGAGTVSAEDKVANDTTAAQTAVGVDTGQDQ 59
DB 1	MEKNVRFKMHKVKGRWVLSVASATMLASALGASVASADTDTASDDSNQTVV---TG-DQ 56
QY 60	ATTNDA--NTNTTDTTDDQSANTNODQAGSDQSNNOQAKODTANTDRNQADNSQTNN 117
DB 57	TNNQATDQTSIAATATSEQSASTD---AATDQASAAEQTQTTASTD---TAAQTITNAN 111
QY 118	QATDOATSPATDG-TSVQRDRDAANVATADQEGOTAPSEQKSAALSNDVKLIDGKY 176
DB 112	EAKWPTENENOGPTDEMLAEAKNVATA---ESDSIPSLAK-----MSNVKQVDGKY 163
QY 177	VOADSGYKKNFALTIVNGQMLYFDSOTGALSSTSTVSFSGGTTNLVDD---PSSHKA 233
DB 164	YDQDGNVKNFPAVSVDGKIYYFD-ETGAYKDTSKVDADKSSSAVSQNAATIFAANNR 222
QY 234	TAKSFELVNGYLTANSWYRPAGILRNGQTVASNSNDLRPVLMSWPKDQTVAYVNY 293
DB 223	SAKNFEADVNYLTADSWYRPKSLDKGKTWTSKGDDFRPLLMWMPDTEKRTNYNN 282
QY 294	KYLSANETEVNTEVSQDILNKEAQSIQTKEQKITSNDSNWLRTAMEAFVAAQPKW 353
DB 283	KVVGIDKT-YTAETSQADLTAAELVQARIEQKITSNNYKWLREASIAFVKTQ 341
QY 354	TENFNKGDLHOGGALLYTN-SOLTFWANSDYRLNRTPTQDQGTK-----KYTEGEG 408
DB 342	SEK-PYDDHLQNGALLFDNQDLDTPDQSNRYLLNRTPTNQTGSLDSRFTYNPND 400
QY 409	ELLNSNDVNSNPVQAEQLNOLHYLMWGDIVMGDKANFDGVRDADVNNADLLQVY 468
DB 401	DFLLANDVDNSNPVQAEQLNHLWLLNLFNFSIYANDADANFDSIRVDAEDNDVDA 460
QY 469	SNYFKDNYKVTDSEANALAHISILEAWSLNDQYNEDTNGTALSIDNSSRLTSLAV 528
DB 461	SDYLKAAAGIDKNNKANNHVSIVEASNDPTFLHDDGDNLMNDNKRFLSMLWSLAK 520
QY 529	PGQRIDLNSLISEVKNKERANTAYGDTIPTYSFVRAHDSVQTVIAKIVKEKIDTNS 588
DB 521	TDVRSGLNPLIHNSL-VDREVDDREVETVPSYSFARAHDSVQDIIRDIKAEINP 579
QY 589	YTFLLDOLKDAEKINEDMAKVNKTYTHYNI PAAYALLSNMESVPRVYVGLYTDGQY 648
DB 580	YSFTQEEIIDQAKFINEDLKSKDKYTHYNVPLSYTLNLTNKGSIPIRVYVGMDF 639
QY 649	MAKKSPPYDAIATMLQGRITAYVSGQSEEVHKVNGNNOILSSVRYGQDLMSADDT 708
DB 640	MANKTVNYDAIESLLKARKMYVAGQAMQNYQI-NGEILTSVRYGKGLKQSD-KG 697
QY 709	SRTSGVLTVLNSDNLDLGGDSLTVNMGRAHANQAYRPLILGTCKGQVQSLKDS 765
DB 698	TRTSGVGVVGNQPNFPLDGLKVVALMGAHAHQBYRALMVSTKDGCVATYATDAD 757
QY 766	IVKYTDANGLNFTADDIKGYSTVDMSGVLAWVPVGAKGODVRAADTNQADKGS 825

DB 158	LVKRTDENGILYFLNDDLLKGVANPQVSGFLQWVPVGAADDQDIRVAASDASTD 817
QY 826	TSALDSQVIYVEGFSNFQDPANNDADYTNKKIAENADFFKLGKLTSPFMAPOYVSAT 885
DB 818	QDAAMDSTRMFEFSGFNSFQSFATKEEBEYTNVVIANNVDFKFSWGIITDFEMAPOY 877
QY 886	FLDSIIQNGYAFSDRYDLAMSKNNKYGSKDDLALNALKALHANGIOAIAIDWPDQ 945
DB 878	FQDSVIQNGYAFTRDRIYDLGMSKANKYGTADQLVKAIKALHAKGLKLMADWPDQ 937
QY 946	BEVTTAKTNSYGNPTFDAYINNALYATNTKSSGSDYQAOYCGAFDLBELKAKY 1005
DB 938	QEVTVTRTDKFGKPIAGSQINHSLYVTDTKSSGDDYQAKYGGAFDLDELK 997
QY 1006	MISTGKPIDPSTKIKQWEAKYFNGTNGVLGKAGYVLSDDATGKYKTYTNVNGD 1065
DB 998	QMSTGQAIDPSPVKIKQMSAKYFNGSNILGRGADYVLSQVSNKYFNVAASDT 1057
QY 1066	DONAKTGFYYDGTGMAYYST-SGNKAVNSFYEGGHYYVFDKDHMTVTSYKAED 1124
DB 1058	KVESGIRYDGKGYIYNSSATGQVKASFITEAGNLIFYFGKDGVMVTGA-OT 1115
QY 1125	FLPNGIQWRDAIYODAOQNSYYYGRGILYKGDNNYFPVDPNNANKTVFRYFD 1184
DB 1116	FLENGALRNTIYTDAQCNHYYANDGKRYENENGYQQFGND-----WRYFK 1168
QY 1185	GYRNYGQTYTFDENGFOAKQOLL-TDDKTHYFPEDNGAMAKNFV-NVGDDW 1242
DB 1169	GLTVDGNVQYFDKDGVOAKDKIIVTRDGKRVYFPDQHNNAVNTFIADKTHW 1228
QY 1243	GNVAKGQYPVNNOILYFNPETGVQVKGFITDAOGRTSYVDANSALSSGFF 1302
DB 1229	GVAVTGAQTVGKQKLYFE-ANGQQVKGDFVTSDEGKLYFYDVDSDGMMW 1287
QY 1303	YYAENGYVYKGFQVAENQDQWYFDQTTGKQAKGAAGVDRDLYFNPDSPGV 1362
DB 1288	FYLGK-----DGAAVTGAQTIROKLYFKA-NGQQVKGDIVK 1323
QY 1363	DESNTSPYHNGDKVVGFFTTGNNAWYADNNG-----NLVKG----- 1406
DB 1324	GTGDKIRYDAKSGEQVFNKTKAADGKTYVIGNDGVAVDPSVVRKGOTFKDAS 1383
QY 1407	IDGK-----WY---HFDEV---TGOAKGAALVNGQOLYFVDVDSGIOVK 1454
DB 1384	LKGQVLTSGWYETANHDWVYIQSKALTGQBTINGQHLYPKKD-GHQVKQLV 1442
QY 1455	TSYDVNSGDKKVGFFTTGDNWY----- 1480
DB 1443	VRYDANSQDQAFNKSVTNGKTYFFGNDGTAQTAGNPKGOTFKDGSDIRFY 1502
QY 1481	-----ADGQ-----GNLAKRKSIDNQLYFPDPATGKQVKQLV-SIDGR 1526
DB 1503	SGWYSNAQGOQLYVKNKVLTLGQTVGSQRVYFD-ENGIOAKGAVRTSDKIRY 1561
QY 1527	SGNMAKPFVRIGDQWIFYGNDGA 1550
DB 1562	SGSMITNQWKEVNGRYFFGNDGA 1585
RESULT 14	
Q55265	STRSL
ID	Q55265_STRSL PRELIMINARY; PRT; 1577 AA.
AC	Q55265;
DT	01-NOV-1996 (TremBLrel. 01, Created)
DT	01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT	01-JUN-2003 (TremBLrel. 24, Last annotation update)
DE	Glucosyltransferase precursor.
GN	Name=gtfm;
OS	Streptococcus salivarius.
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC	Streptococcus.
OX	NCBI_TaxID=1304;
RN	[1]

RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=95122197; PubMed=7822030;
RA Simpson C.L., Giffard P.M., Jacques N.A.;
RT "Streptococcus salivarius ATCC 25975 possesses at least two genes
coding for primer-independent glucosyltransferases.";
RL Infect. Immun. 63:609-621(1995).
DR EMBL: L35928; AAC41413.1; -; Genomic_DNA.
DR FIR; T30858; T30858.
DR HSP; P06653; 1H8G.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR004829; C:surface antigen.
DR InterPro; IPR002479; C:binding.
DR InterPro; IPR003318; G:glyco_hydro_70.
DR Pfam; PF01473; C:binding 1; 4.
DR Pfam; PF02324; G:glyco_hydro_70; 1.
DR ProDom; PD153432; C:surface_antigen; 1.
KW Signal; Transferase.
FT SIGNAL 1 38 Potential.
FT CHAIN 39 1577 Glucosyltransferase.
SQ SEQUENCE 1577 AA; 175290 MW; 3EFB989A7D3A7BF3 CRC64;
Query Match 41.6%; Score 3429.5; DB 2; Length 1577;
Best Local Similarity 44.7%; Pred. No. 3.2e-141;
Matches 738; Conservative 250; Mismatches 471; Indels 193; Gaps 40;
QY 1 MEKKLHYKLHKVGHVTTIAVASIGLVSLVGAGTVAEDKVAANDTTA----- 47
DB 1 MENKVRFLKHKVKNVTTIGVITLSMVALAG-GSLLAQCKVEADETSAPNGDGLQOLSED 59
QY 48 -----QATVGVDTGQDQATNDANTNTDTADQSANTNQDQAGSDQSNQDQAKQDT 101
DB 60 GTASLVTMTTTEQASQAASAVATASVSHETSFOAATSQVSEATAQATSPVASQEV 119
QY 102 ANTDNRQADNSQD-----NNQATDQATSPATDGTQVRD-----AANVAT-A 144
DB 120 AVSSQTSQSGQETQTEQVSQGTSTQVAGQTSQASTPSVTEQARPRVLTAAPAATRA 179
QY 145 AD-----QEGQ- 150
DB 180 ADSTIRANRNTNITITAGTTPNVTITGNTPKPNVTVTPNGTRPNVTIVTQNPQ 239
QY 151 -----TAPSEQKSA-----NKLIDGKYVVOAGSGYKKNFAITVNGQM 195
DB 240 NKEVPQSPQSPQPNQPSLDYKPVASNLKTIDGQYVVE-NGVVKNAIAELDGL 298
QY 196 LYFQSDTGAL--SSSTSYFSQGTTLNVDPSHNKAYDSTAKSPFLVNGYLTANSWRP 253
DB 299 YTFD-ETGAMVDQSKPLRADAI PNNSI--YAVYQAYDTSKSPFHLNFLTADSWYRP 355
QY 254 AGILRNGQTWEASNENDLRPVLMSWMPDKDTQAVYVYNNKYLSANETEVTNETSOVDLN 313
DB 356 KQLKDGKNWTASTKEDRPLMTWHPDKVTQVYLYNYSQQGFGNKYITTDMS-YDLA 414
QY 314 KEAQSIQTKIEOKITSDNSTQWLRTAMEAFVAAQPKNNMSTB-NFNKG-DHLQGGALLYT 371
DB 415 AAAETVQRIEGRIGREGNTTWLRQLMSDFIKTQPCWNSSESDNLLVGKDHLLQGGALTF 474
QY 372 NSDLTPWANSYRLNRPQTQDGKTKYFTEGEGGYFELLSNDVNSPVPVQAOLNQL 431
DB 475 NNSATSHANSDFRLNRPNTQGTGRKYHIDRSNGGYELLANDINDNSNPAVQAOLNWL 534
QY 432 HYLNMWGDIVMGDKDANPDGVRVADVNVNADLLQVYSNYKDYKVTDSANALAHISI 491
DB 535 HYIMIGSILGNPDSPANDGVRIDAVDNDVADLLQIASDYFKEKRVADNEANAHLISI 594
QY 492 LEAMSLNDQNYENDTNGTALSIDNSRLTSLAVLTQPCQRIDLSNLISSEYKNERANDT 551
DB 595 LEAMSYNDHQYNKQTKGAQLSIDNPLRETLTFLRKSNYRGSLSERVITNSLN-NRSE 653
QY 552 AYGDITPYFSVRAHDSQVTVIAKIVKEKIDTSDGVTFTLDQLDKAFKYNEMAKVN 611
DB 654 KHTPRDANYIFVRAHDSQVAVLANIISKQINPKPTDGTFTMDLQKAFEINYADIADK 713

RESULT 15

GTFC_STRMU

ID GTFC_STRMU STANDARD; PRT: 1455 AA
AC P13470; O69382; O69385; O69386; O69391; O69397; P05427;
DT 01-NOV-1988 (Rel. 09, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)

QY 612 KTYTHYNI PAAYALLSNMESPRVYVYGLYTDGQYMAKSPYDAIATMLOCRYAVS 671
DB 714 KKTYQYNI PAAYATMLTNKDSITRVYVYGLYTDGQYMAEKSPYYNAIDALLARIKYVA 773
QY 672 GGQSEVHKVNGNQNQLSSRVYVYVQDLSADDTGDTLSTRTSGLVTLVSDNPDLGL-GDS 730
DB 774 GGQDMKVTKLNG-YEIMSSVRYKGAEEANQL-GTAEITRNOQMLVLTANRPMKLGANDR 831
QY 731 LTVNMGRAHANQAYRPLIILGTGQVOSYKLDSD--TNIYKYTDPANGLTFADDIKGYST 788
DB 832 LVNMGAAHKNQAYRPLLSKSTGLATYKDSVDPAGLVRYTDPNQNGLTFADDIAGHST 891
QY 789 VMSGYLA VVVPVGAQGDQVRVAADTNQADKSLKTSAAALDSQVIYEGFSNFQFPANN 848
DB 892 VEVSGYLA VVVPVGAASENQDARTKASTKKGE-QVFESSAALDSQVIYEGFSNFQFVK 950
QY 849 DADYTNKIAENADPFKLGITSFEMAPQYVVSATGSDFLDSIIQNGYAFSDRYDLAMSKN 908
DB 951 PSQYTNRVIAQNAKLFKEWGITSFEPAPQYVSSQDGTFLDITENGYPEDRYDIAMSKN 1010
QY 909 NKYGSKDDLANALKALHANGIOAIAADWVDPDIYQLPGEVVTAKRTNSYGNPTFDAYINN 968
DB 1011 NKYGSUKDLMDALRALHAEGISAIADWVDPDIYNLPGEVVTASRTNSYGTPRPNABIYN 1070
QY 969 ALYATNTKSSGSDYQAQYGGAFDELKAKYPMFTVMNISTGKPIDPSTKIKOWEAKYFN 1028
DB 1071 SLYAAKTRTFGNDFOGKYGGAFDELKAKYPAIFERVOISNGRKLITNEKITOWSAKYFN 1130
QY 1029 GTNVLKGGAGYVLSDDATGKYFTVNENSGDFLPASFTGQNAKTFYFYDGTGMAYSTSGN 1088
DB 1131 GSNIGTGARYVLDNATNQYFSVKAQTFLPKQMT--EITGSGFRVRVDDVQYLSIGGY 1188
QY 1089 KAVNSFIYEGGH-YYEFKDGCHMTGSKYKAENDYVYFLPNGIOMRDALYQDAQNSYYY 1147
DB 1189 LAKNTFIQVGAQWYFYFDKNGNMVTGE-QVIDGKYFFLDNGLQLRHVLRQSGDGHVYY 1247
QY 1148 GRGTGI-LYKGDNWPYFVDPNNANKTVFRYFDANNVMAIGYRNMYGQTYVYFDE-NGFQAKG 1205
DB 1248 DPKGVQAFNG--FYDFAGP---RQDVRYFDGNGQMYRGLHDMYGTTFYFDEKGTGLQAKD 1301
QY 1206 QLLT-DDKGTHYFDEBDNGAMAKNFVNVGDD--WYMDGNGNAVKGQYVNNQILYFNPE 1262
DB 1302 KFIREADGRTRYFIPDTGNLAVNREAFQNPENKAWYLDNSGYAVTGLQTINGKQYFDNE 1361
QY 1263 TGVQVKGQFIITDAQRTSYVDANSALXSSGFTTNGSDWYVAENGYYVYKGFQVAENQD 1322
DB 1362 -GRQVKGHFVT-INNQRYFLDGDGSEIAPSRFVTEN-----N 1396
QY 1323 QWYVYFDQTTGQAKGAAGVDRDLVFNPDGSGVQVKGDPATDESGNTSFYHGDNGDKVVG 1382
DB 1397 KWYVYD-GNGKLVKGAQVINGNHYYFNNDYS-QVKGAWA-----NGRYVDGSGQAVSNQ 1449
QY 1383 FPTTGNNAWYADNNGNLVKGFQEI DGKWHYFDEVTGQQAAGALVNGQOLYFDVDSGIG 1442
DB 1450 FTQIAANQWYLVNQDGHKVTGLGNINNKVYVYG-----SNGAQ 1487
QY 1443 VKGDFVTQCGNTSYVDVNSGDKKYNNGFTTGDNNAWYADGQGNLAKGRKSTDNQDLYFD 1502
DB 1488 VKGKLLT-VQGGKCYFDAHTGCVVYVNRFEAARGCWYFNSAGQAVTGOVINGKQLYFD 1546
QY 1503 PATGQVKGQOLVSIDGRNYYFDPGSGNNMAKNR 1534
DB 1547 -GSGRQVAGRYVYVGGKRLFCDAKTGELRQR 1577

QY 57 QOATINDANTWT-TDITADOSANTNQOAG-SDQSNQOQAKQDTANT-----DRNQ 108
Db 56 ---VTTSEAAKELTATDSTATSATSQPTATVTDNVSTTQNTTANTANTANFDVKTPTT 112
QY 109 ADNSQTDNNO---ATDOATS-----PATDGTSVORRDAANVATAADQEG---OTAP 153
Db 113 SQSKTDNSDKIATSKAVNRLTATGKFPANNNTAHSRTVTDKVIPIKPKGLKQPS 172
QY 154 SQEKSAAI-SLDNVKLIIDGKYVVOAGDSYKKNFAITVNGQWLYFDSOTGALSSTSTYS 212
Db 173 LSQDDIAALGNVKNIRKVGKYYYKEDGTQKNYALNINGKTFPD-ETGALSNNLTLS 231
QY 213 FQGITN--LVDDPSHNKAYDSTAKSPFLVNGYLTANSWYRPAIILRNGQOTWEASNEND 270
Db 232 KXGNJNNNDNTNSFAQYNQVYSTDAANFEHVDHYLTAEWSYRPKYILKDKGTWQSTEXD 291
QY 271 LRPVLMSWPDKDTQVAYVYNNKXLSANETEVNETSOVDLNKEAOSIOTKLEOKITSD 330
Db 292 FPLLTMTWPDQETQRYVYNNQAQGIHQTYNT-ATSPQLNLAAQTIOTKLEKITAE 350
QY 331 NSTOMLRTAMEAFVAAQPKWNMSTEN-FNKGDHLOCGALLYT-NSDLTPWANSDYRLNLR 388
Db 351 KTNMLRQITAFVKTQSAWNSDSEKFPD--DHLQKALLYSNNSKLTQANSNYRILNR 408
QY 389 TPTQODGK--KYFTREGGEGYEFLLSNDVNSNPVQAEQNLQHLHLMWGDIVMGDKD 446
Db 409 TPTNQTGKDPRTADRTTGGYEFLLANDVNSNPVQAEQNLWHLFLNFGNIYANDPD 468
QY 447 ANFDGVRVDAVNVAADLLOVSNYFKDNYKYVTDSEANALAHISTLEAMSNDNQNEBT 506
Db 469 ANFDSIRVDAVNDAADLQIAGDYLAAGKHNDKAAANDHLSLEAWSYNDTPYLHDD 528
QY 507 NCTALSIDNSRLTSLAVITKPGQORIDLNSLISESVNKERANDTAYGDTIPTYSFVRAH 566
Db 529 GDNMINMDNRLSLYLAKPLNQBSGNPLITNSL-VNRDNDNAETAAVPSYFIRAH 587
QY 567 DSEVQTVIAKIVKEIKDITNSDGYTFTLDQKDAFIYNEDMAKVNKYTHYINIPAAAYALL 626
Db 588 DSEVQDLIRIIRAEINPNVGVSYFTMBEIKKAFIYNKDLLATEKKYTHYNTALSYALL 647
QY 627 LSNMESPRVYVYDLYTDDGOYMAKSPYDYAIAIATMLQRIAYVSGGSEEVHKVNGNQ 686
Db 648 LTNKSSPRVYVYDLYTDDGOYMAHKTINYEAETLLKARIKYVSGGQAMRNQV--GNSE 706
QY 687 ILSVRYGODLMSADDTQGTDLSTRSGLVTLVSNPNLTL-GGDSLTVMGRAHANQAYR 745
Db 707 IITSVRYGKALKATDT-GDRTTRSGVAVIEGNPFLKASDRVNVNMGAAHKNQAYR 765
QY 746 PILGTDGVSQSYLXDS--TNIVKYTDANGNLTFTADDIKYSTVDMSCYLAIVVPVGA 804
Db 766 PLLLTDDNGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKGYANPQVSGYLGWVPVGA 825
QY 805 DQDVRVAADTNQKADGSKLTSALDSQVIYEGFSNFPQDPAANDADYTNKKIAENADPF 864
Db 826 ADQDVRVAASTAPSDGKSVHQNAALDSRMFEGFSNFOAFATKKEEYTNVVIKKNVDF 885
QY 865 KKLGITSEMAYQYVYATDGSFLDSIIONGYAFSDRYDLMSKNNKYGSKDDLANALKAL 924
Db 886 AEWGVTDFEMAPQYVSSDGSFLDSVIQNGYAFTRDYDLGISKPNKIGTADDDLVKAIKAL 945
QY 925 HANGIAIADWVPDQIYQIPGEVVTAKRTNSYGNPTFDAYINNALYATNTKSSGSDYQA 984
Db 946 HSKGIKVMADWVPDQYALPEKEVVTATRVDKYGPVAGSQIKNTLYVVDGSSGKQQA 1005
QY 985 QYGGAFLELAKYQYVDMFTVMNISTGKPIDPSTKIQWEAKYFNGTIVLKGAGYVLSDD 1044
Db 1006 KYGGAFLBELOAKYPELFAKQISTGVPMDPSVKIKQMSAKYFNGTIVLGRGAGYVLDQ 1065
QY 1045 ATGKYFTVNEGDPLPASFT----GDQNAKTGFYVDGTGMAYYSTSGNKAVNSFIYEGGH 1100
Db 1066 ATNTYFSLVSONTFLPKSLVNPNNHGTSSSVGLVFDGKYVYVYSTSGNQAKNAFISLGN 1125
QY 1101 YYYFDKDGHMVTGSKAEDGNDYFPLNGIQMRDAIYQDAQNSYYYGRTGILYK---- 1155

Db 1126 WYFDNNGYMWVGA-QSINGANYFYFLSNGITQLRNLAYDNGNKVLSYYGNDGREYNGYL 1184
QY 1156 -GDNWYFPVDPNNANKTVPRYFDANNVMAIGYRNMYGQTYYPDENGFQAKGOLLTDDKG- 1213
Db 1185 FQOQW-----RYFQ-NGIMAVGLTRIIGHAVQVFPASGFQAKGQFITTADGK 1229
QY 1214 THYFDEDNAMAKNFV-NVGGDDWYMDGNGNAVKQGYFVNNQILYFNBPETGVQVKGQFI 1272
Db 1230 LRYFDRDSNQISNRFPVRSKGWFLFDHNGVAVTGTVTFNGQRLYFKP-NGVQAKGEFI 1288
QY 1273 TDAQGTSTYDANSALKSSGFTPNGSDWYYAENGYVYKGFQKVAENQDQWYFPDQTTG 1332
Db 1289 RDADGHLRYDPSNGNEVRNF-----VRNSKGWFLFDH-NG 1325
QY 1333 KQAKGAAKVDGRDLVFNPDGQVQVKGDFATDSSGNTSFYHGGNDGDKVGGFFTTGNNAVY 1392
Db 1326 IAVTGRVNVGQRLYFK-SNGVQAKGELLITERKGRIKYYDPNSGNEVRNRYVRTSSGNWY 1384
QY 1393 YADNNGNLVKGFEIDGKWYHDFEVTGQQAAGAAALVNGQQLYFVDVDSGIQVKGDFVTD-G 1451
Db 1385 YFGNDGYALIGWHVVEGRVYFDE-----NGVYRYASHDQORNHWDYDVRDPG 1432
QY 1452 QGNTSYDVNSGDKXVNGFFTTGDN 1476
Db 1433 RGSASSAVRFRHSR---NGFF---DN 1451

Search completed: February 11, 2006, 19:38:42
Job time : 222.688 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 18:56:56 ; Search time 135.868 Seconds
(without alignments)
4414.224 Million cell updates/sec

Title: US-10-797-821-39
Perfect score: 7230
Sequence: 1 MEKNRYKLHKVKKQWVAIG.....WKNVDGNWYFNRGLATRW 1365

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*
9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7230	100.0	1365	7	ADD93659 Streptococcus
2	7230	100.0	1365	9	ADK37282 Streptococcus
3	3174	43.9	1430	5	AU98029 S. mutans
4	3174	43.9	1430	7	ADD93656 Streptococcus
5	3174	43.9	1430	9	ADK37279 Streptococcus
6	3169	43.8	1430	5	AU98043 S. mutans
7	3168	43.8	1430	5	AU98042 S. mutans
8	3168	43.8	1430	5	AU98041 S. mutans
9	3163	43.7	1430	5	AU98044 S. mutans
10	3163	43.7	1430	5	AU98045 S. mutans
11	3080.5	42.6	1554	9	ADD93658 Streptococcus
12	3080.5	42.6	1554	9	ADK37281 Streptococcus
13	3031	41.9	1590	7	ADD93657 Streptococcus
14	3031	41.9	1590	9	ADK37280 Streptococcus
15	3026.5	41.9	1577	2	AAR91047 Alpha-D-g
16	3023	41.8	1592	2	AAR32925 Glucosyl
17	3001	41.5	1476	5	AU979284 Streptococcus
18	2973.5	41.1	1475	5	AU98038 S. mutans
19	2963.5	41.0	1475	5	AU98037 S. mutans
20	2957.5	40.9	1475	5	AU98036 S. mutans
21	2956.5	40.9	1475	5	AU98035 S. mutans
22	2956.5	40.9	1475	5	AU98034 S. mutans
23	2951.5	40.8	1475	5	AU98033 S. mutans
24	2951.5	40.8	1475	5	AU98032 S. mutans

25	2950.5	40.8	1475	5	AU98031 S. mutans
26	2949.5	40.8	1475	5	AU98040 S. mutans
27	2946.5	40.8	1475	5	AU98030 S. mutans
28	2945.5	40.7	1475	5	AU98027 S. mutans
29	2945.5	40.7	1475	7	ADD93654 Streptococcus
30	2945.5	40.7	1475	9	ADK37277 Streptococcus
31	2939	40.7	1499	7	ADC54806 Protein S
32	2929.5	40.5	1475	5	AU98039 S. mutans
33	2893.5	40.0	1375	5	AU98028 S. mutans
34	2893.5	40.0	1375	5	AU979288 Streptococcus
35	2893.5	40.0	1375	7	ADD93655 Streptococcus
36	2893.5	40.0	1375	9	ADK37278 Streptococcus
37	2891.5	40.0	2835	5	ABB98574 Dextran s
38	2891.5	40.0	2835	6	ABR55594 Amino aci
39	2804	38.8	1518	7	ADD93660 Streptococcus
40	2804	38.8	1518	9	ADK37283 Streptococcus
41	2680.5	37.1	1497	6	ABR63234 Glucansuc
42	2661.5	36.8	1477	9	ADY72733 Mutant de
43	2656.5	36.7	1477	9	ADY72696 Mutant de
44	2653.5	36.7	1477	9	ADY72732 Mutant de
45	2640	36.5	1527	7	ADC54807 Leuconost

ALIGNMENTS

RESULT 1

ADD93659	ADD93659 standard; protein; 1365 AA.
XX	XX
AC	ADD93659;
XX	XX
DT	29-JAN-2004 (first entry)
XX	XX
DE	Streptococcus downei glucosyltransferase-S.
XX	XX
KW	Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.
XX	XX
OS	Streptococcus downei.
XX	XX
PN	WO2003075845-A2.
XX	XX
PD	18-SEP-2003.
XX	XX
PF	07-MAR-2003; 2003WO-US006962.
XX	XX
PR	07-MAR-2002; 2002US-0363209P.
PR	08-AUG-2002; 2002US-0402483P.
XX	XX
PA	(FORSYTH INST.
PI	Smith DJ, Taubman MA;
XX	XX
DR	WPI; 2003-845091/78.
XX	XX
PT	Composition useful as vaccines for dental caries comprises a fragment of a glucan binding protein-B binding to a major histocompatibility complex class II protein.
XX	XX
PS	Claim 16; Page 15-16; 49pp; English.
XX	XX
CC	The present sequence is the protein sequence of Streptococcus downei glucosyltransferase-S (GTF-S). Peptide fragments of GTF-S, especially from the catalytic domain of the polypeptide, can be used in immunogenic compositions and subunit vaccines for dental caries. These compositions comprise a major histocompatibility complex (MHC) class II protein-binding peptide from S. mutans glucan binding protein-B (GbpB) covalently linked with a peptide fragment of a streptococcal glucosyltransferase. The compositions are used in a claimed method of eliciting production of an antibody in a mammal. Dieptopic or multiepitopic polypeptides can be prepared synthetically or by recombinant DNA technology. Antibodies raised against MHC class II binding fragments of GbpB can be used in passive immunisation.

CC immunizing mammals against dental caries. This sequence corresponds to a
CC Streptococcus downei GbpB protein of the invention.
XX
SQ Sequence 1365 AA;

Query Match	100.0%;	Score 7230;	DB 9;	Length 1365;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1365;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MEKNRYKLVKVKQWVAIGVTTVLSFLAGGQVVAADTNNDGTSVQVNMKVPSPKPD	60	
DB	1	MEKNRYKLVKVKQWVAIGVTTVLSFLAGGQVVAADTNNDGTSVQVNMKVPSPKPD	60	
QY	61	AOAONGOLAQAMFKAANQADQATTSQVSPATDGRVDNQVTPAANQPAANQVANQVANPAT	120	
DB	61	AOAONGOLAQAMFKAANQADQATTSQVSPATDGRVDNQVTPAANQPAANQVANQVANPAT	120	
QY	121	DAGALNRQSAADTSTGKAVPQTSQPGHLETVGKTYVDANGQRLKNSYMWIDGKTYT	180	
DB	121	DAGALNRQSAADTSTGKAVPQTSQPGHLETVGKTYVDANGQRLKNSYMWIDGKTYT	180	
QY	181	FDGQTGEAOTDLPKTGOAQNDVPDSYQANNOQAYSNEASSFETVDNYLTADSWYRPRKIL	240	
DB	181	FDGQTGEAOTDLPKTGOAQNDVPDSYQANNOQAYSNEASSFETVDNYLTADSWYRPRKIL	240	
QY	241	KNGSQWQASSEGDLRPIILMTWPDAAATKAAYANFWAKEGLISGYSYRQNSANLDAATQNIQ	300	
DB	241	KNGSQWQASSEGDLRPIILMTWPDAAATKAAYANFWAKEGLISGYSYRQNSANLDAATQNIQ	300	
QY	301	SAIEKKIASEGTNWLRKMSQVFKSQNQSIASENETVYPNODHMQGALLFSPNSKQTE	360	
DB	301	SAIEKKIASEGTNWLRKMSQVFKSQNQSIASENETVYPNODHMQGALLFSPNSKQTE	360	
QY	361	HANSWRLLNRNPTFOTGKQKPTTNYAGYELLANDVNSNPVQAEOLNHLHYLMWNG	420	
DB	361	HANSWRLLNRNPTFOTGKQKPTTNYAGYELLANDVNSNPVQAEOLNHLHYLMWNG	420	
QY	421	DIVMGDKANFDGVRVDAVDNVNADLLQIQDYKAKYGTQDQEKNAIDHLSILEAWSN	480	
DB	421	DIVMGDKANFDGVRVDAVDNVNADLLQIQDYKAKYGTQDQEKNAIDHLSILEAWSN	480	
QY	481	DNDYVKDQNNFSLISINDQSGMLKAFGVASAYRGNLSNATAGLNKRSANPDSVPVNY	540	
DB	481	DNDYVKDQNNFSLISINDQSGMLKAFGVASAYRGNLSNATAGLNKRSANPDSVPVNY	540	
QY	541	VFIRAHDSVQTRIAKIIREKLGKTNADGLTNLTDLANKAFDIYNQDMNATDKVYYPNN	600	
DB	541	VFIRAHDSVQTRIAKIIREKLGKTNADGLTNLTDLANKAFDIYNQDMNATDKVYYPNN	600	
QY	601	LPWAYAWMLQNDTTRVYVYGDVMTDNGQYMATKTPFYNAIETLLKGRIKYVAGQAVSY	660	
DB	601	LPWAYAWMLQNDTTRVYVYGDVMTDNGQYMATKTPFYNAIETLLKGRIKYVAGQAVSY	660	
QY	661	KQDWSGILTSVRYGKANSASDAGTETRNQSMALLINRNFRAYRNLTLNMGAAHKS	720	
DB	661	KQDWSGILTSVRYGKANSASDAGTETRNQSMALLINRNFRAYRNLTLNMGAAHKS	720	
QY	721	QAYRPLLSTKDGIAIYLNDSVDVSRQYKTDQGNLSFASSELQSVANAQVSGMIQWVY	780	
DB	721	QAYRPLLSTKDGIAIYLNDSVDVSRQYKTDQGNLSFASSELQSVANAQVSGMIQWVY	780	
QY	781	PVGAADNQDVRTSPSTQATKQGNIIHQSDALDSQVIYEGFSNFQAPQSPDOYTNVIAK	840	
DB	781	PVGAADNQDVRTSPSTQATKQGNIIHQSDALDSQVIYEGFSNFQAPQSPDOYTNVIAK	840	
QY	841	NGDLFKSWGITQFEMAPQVVSSEDGTFLLDSVILNGYAFSDRYDLAMSKNKYGSQDLAN	900	
DB	841	NGDLFKSWGITQFEMAPQVVSSEDGTFLLDSVILNGYAFSDRYDLAMSKNKYGSQDLAN	900	
QY	901	AIKGLQSAQIKVLSLVNQLNLPFGKEVVTATRVNVOYQAKSGATINKTPVANTRSYG	960	
DB	901	AIKGLQSAQIKVLSLVNQLNLPFGKEVVTATRVNVOYQAKSGATINKTPVANTRSYG	960	

QY	961	DYQEQYGGKFLDLDLQKLYPRLFSTKQISTGKPIDPSVKITNWSAKYFNNGSILGRGAKYV	1020	
DB	961	DYQEQYGGKFLDLDLQKLYPRLFSTKQISTGKPIDPSVKITNWSAKYFNNGSILGRGAKYV	1020	
QY	1021	LSEGNKYLNLADGKLFPLTVLNTYQPOQVPSANGFISKNGGIIHYLDKNGQEVKNRFEIS	1080	
DB	1021	LSEGNKYLNLADGKLFPLTVLNTYQPOQVPSANGFISKNGGIIHYLDKNGQEVKNRFEIS	1080	
QY	1081	GSWYFSDSGKMATGKTKIGNDTYLPMPNGKQJKEGVWYDGGKAYYDDNNGRTWTNKGfV	1140	
DB	1081	GSWYFSDSGKMATGKTKIGNDTYLPMPNGKQJKEGVWYDGGKAYYDDNNGRTWTNKGfV	1140	
QY	1141	EFVRDQDKWRYFNCGDTTIAIGLVSLDNRTLYFDAYGYQVKGQTVTINGKSYTFDADQGD	1200	
DB	1141	EFVRDQDKWRYFNCGDTTIAIGLVSLDNRTLYFDAYGYQVKGQTVTINGKSYTFDADQGD	1200	
QY	1201	LVQTDNANPAPOQAGWKLKLDNNGYRKGQGLLTGEQITIDGQVFFQDNGVQVKGGTAT	1260	
DB	1201	LVQTDNANPAPOQAGWKLKLDNNGYRKGQGLLTGEQITIDGQVFFQDNGVQVKGGTAT	1260	
QY	1261	DASGVLFYDRDQGHQVKGWYSTDNDNMYVNESQVLTGLQITIDGQTVYFDDKGIQAK	1320	
DB	1261	DASGVLFYDRDQGHQVKGWYSTDNDNMYVNESQVLTGLQITIDGQTVYFDDKGIQAK	1320	
QY	1321	GKAVMDENGLRYFDADSGNMLRDRWKNVYDGNWYFNRRNGLATRW	1365	
DB	1321	GKAVMDENGLRYFDADSGNMLRDRWKNVYDGNWYFNRRNGLATRW	1365	
RESULT 3				
AAU98029				
ID	AAU98029 standard; protein; 1430 AA.			
AC	AAU98029;			
XX	27-AUG-2002 (first entry)			
DE	S. mutans glucosyltransferase GTFD.			
KW	Glucosyltransferase; GTFD; transgenic plant; paper sizing;			
KW	coating composition; glucan; starch; latex; thermoplastic molecule;			
XX	amyloplast; vacuole; paper manufacture.			
OS	Streptococcus mutans.			
PN	US2002031826-A1.			
XX	14-MAR-2002.			
XX	19-DEC-2000; 2000US-00740274.			
PR	07-JUN-1995; 95US-00478704.			
PR	07-JUN-1995; 95US-00482711.			
PR	16-JUN-1995; 95US-00485243.			
PR	16-JAN-1998; 98US-00007999.			
PR	16-JAN-1998; 98US-00008172.			
PR	20-JAN-1998; 98US-00009620.			
PR	11-DEC-1998; 98US-00210361.			
XX	(NICH/) NICHOLS S E.			
XX	Nichols SE;			
XX	WPI; 2002-414332/44.			
DR	N-PSDB; ABK52940.			
XX	Glucosyltransferase B or D protein useful for producing a glucan useful			
PT	as substitutes for and additions to modified starch and latexes in paper			
PT	manufacture, comprises mutations in specific positions.			
XX	Disclosure; Page 38-42; 44pp; English.			
XX	The invention an isolated protein comprising a glucosyltransferase (GTF)			

B polypeptide having changes at position from I448V, D457N, D567T, I404L4T, D457N/D567T, D457N/D571K, D567T/D571K/I4014T, I448V/I404L4T/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a GTF D polypeptide having changes at positions from T589D, T589E, N471D, N471D/T589D, and N471D/T589E. Also included are a glucan produced by the GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its complementary polynucleotide, a ribonucleic acid sequence encoding the GTF mutant, an expression cassette comprising the polynucleotide operably linked to a promoter, a vector comprising the expression cassette, host cell introduced with the vector, a transgenic plant comprising the vector, a seed or tuber from the transgenic plant, a paper sizing and/or coating composition comprising a glucan produced in a plant transformed with a gene encoding the mutant GTF, wild type or, starch, a latex, thermoplastic molecule or their combinations or glucan and starch where the glucan is produced in the amyloplast and/or vacuole or a maize line deficient in starch biosynthesis, transformed with a gene encoding a glucosyltransferase B or D enzyme, wild-type or mutant and a paper comprising the glucan (paper sizing/coating agent). The vector is useful for producing a glucan in a plant. The method comprises transforming a plant cell with the vector, growing the plant cell under plant growing conditions to produce a regenerated plant and inducing expression of the polynucleotide for a time sufficient to produce the glucan in the regenerated plant, where the vector contains a transit sequence from ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and chlorophyll AB binding protein to produce a transgenic plant, and glucan is produced in the amyloplast of potato or the vacuole of sugar beet. Glucans are useful as substitutes for and additions to modified starch and latexes in paper manufacture. Unlike prior art techniques, which require input materials that produce chemical effluents, paper manufacture utilising the glucan produced by GTF, which utilises biologically produced input materials, is more cost-effective and environmentally friendly. Moreover, glucans also exhibit thermoplastic properties and impart gloss to the paper during coating step. The present sequence represents GTFD

Sequence 1430 AA:

Sequence 1430 AA;

Query Match 43.9%; Score 3174; DB 5; Length 1430;
Best Local Similarity 47.3%; Pred. No. 5.3e-180;
Matches 677; Conservative 213; Mismatches 436; Indels 104; Gaps 38;

Qy	1	MEKNURYLKHVKYKOWAIVGTT--VTLSFLAGQVVAADTNNDNGTSHV-----QVNKMV	53
Db	1	MEKRYRKNHVKYKHWTVAVASGLITLTGTTTLGSSVSAETEQTGSDKVVTQSEEDKAA	60
Qy	54	PSDPKPFDA----QAQNGLOAQMFKAAQAOADOTATSOVSPATDGRVDNVQVTPAANQPAAN	109
Db	61	SESSQTDAPKTKQAQTEQ--TQAQ--SQANVAD--TSST-----ITKETPSQNTITTOANSDDKT	113
Qy	110	VANQDVANPATDAGALNRQSAADTSTDGKAVP-----QTSQDPG-----H	149
Db	114	VTNTKSEEAQTSEERTKQSEEAQITASSQAULTQAKAELTKQRQTAAQENKNPVDLAAIPN	173
Qy	150	LETVDGKTYVDNANGORLKNYSWIDGKTYFYDCGTGEAQTDLP----KTGQANQDNVPD	205
Db	174	VQIQIDGKYIYIGSDGQPKKNFALTNNKVLFYDKNTG--ALTDTSQVQFKOGLTKFLNN--	229
Qy	206	SYQANNOAYSNEASSFETVDNYLTADSWYRPRKILKNGSQWQASSEGDLRPIILMTWPDA	265
Db	230	DYTPHNQIVNFENTSLEITIDNVYTDADSWYRPKDILKNGKTKWTASSESDLRPLLMSWWPK	289
Qy	266	ATKAAAYANFWAKEGLISG--SYRQNSA--NLDAATQNTQSAIEKKIASEGNTWMLRDMQSQ	322
Db	290	QYQIAYLYMNMQQGLGTGENTYADSSQESLNLAQTQVVKIETKISQTOQTQWLRDIINS	349
Qy	323	FVKSQNQWSIASENETVVPNOCHMOGGALLPSNSKQDTEHANSQWRLLNRRNPTFTQCKQKY	382
Db	350	FVKYTOPNWSQTESDTSAGEKDHLCGGALLYSNSDKTAYANSQRYLLNRRPTTSQTKPKKY	409
Qy	383	FTTNTYA--GYELLANDVNSNPVQAEOLNHLHLYLMNWGDI VMGDKDANFDGVRVDVN	441
Db	410	FEDNSGGYDFLLANDIDNSNPVQAEQLNWLHLYLMNYGSI VANDPEANFDGVRVDVN	469

RESULT 4

RESUL 4
ADD93656

ADD93656
ID ADD93656 standard: protein: 1430 AA.

AC ADD93656:

AC ADD53636;
XX

XX
29-JAN-2004 (LIST ENCLY)

Streptococcus mutans glucosyltransferase-D.
 Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.
 Streptococcus mutans.
 W02003075845-A2.
 18-SEP-2003.
 07-MAR-2003; 2003WO-US006962.
 07-MAR-2002; 2002US-0363209P.
 08-AUG-2002; 2002US-0402483P.
 (FORS-) FORSYTH INST.
 Smith DJ, Taubman MA;
 WPI; 2003-845091/78.
 Composition useful as vaccines for dental caries comprises a fragment of a glucan binding protein-B binding to a major histocompatibility complex class II protein.
 Claim 16; Page 13-14; 49pp; English.
 The present sequence is the protein sequence of Streptococcus mutans glucosyltransferase-D (GTF-D). Peptide fragments of GTF-D, especially from the catalytic domain of the polypeptide, can be used in immunogenic compositions and subunit vaccines for dental caries. These compositions comprise a major histocompatibility complex (MHC) class II protein-binding peptide from S. mutans glucan binding protein-B (GbpB) covalently linked with a peptide fragment of a streptococcal glucosyltransferase. The compositions are used in a claimed method of eliciting production of an antibody in a mammal. Diepitopic or multiepitopic polypeptides can be prepared synthetically or by recombinant DNA technology. Antibodies raised against MHC class II binding fragments of GbpB can be used in passive immunisation.
 Sequence 1430 AA;
 Query Match 43.9%; Score 3174; DB 7; Length 1430;
 Best Local Similarity 47.3%; Pred. No. 5.3e-180;
 Matches 677; Conservative 213; Mismatches 436; Indels 104; Gaps 38
 QY 1 MEKNRYKLHKYKQWVAIGYTT--VTLSFLAGQGVAAATNNNDGTSV-----QVKNW 53
 DB 1 MEKGRYKMKHKYKGMVTVAVASGILITGLTTGLSSVSAETEQTSDKVVTQKSEDDKAA 60
 QY 54 PSDPKFDA-----QAQNGOLAQAMFKAANOADOTATSQVSPATDGRVDNQVTPAANPAAN 109
 DB 61 SSSQTDAPKTKAQTEQ-TQAQ--SQANVAD-TSTS-----ITKTPSQNITTOANSDDKT 113
 QY 110 VANQDVANPATDAGALNRQSAADTSTDGKAVP-----QTSQDPG-----H 149
 DB 114 VINTKSEEAQTEERTKQSEEAQTASSQAALTOAKAELTKORQTAAGKNQVPVDLAAIFN 173
 QY 150 LETVDGKYYVDANGQRKLKNSWIDGTYTFDQGTGEAOTDLP----KTGQANQDNVDP 205
 DB 174 VKQIDGKYYIGSDGQPKNFALTNNKVLVFDKNTG-ALTDTSQYQFKQLTKLNN--- 229
 QY 206 SYQANNOAYSNEASSFETVDNYLTADSWYRPRKILKNGQSMQWASSGDLRLPILMTWMPDA 265
 DB 230 DYTPHNQIVNFENSLFETIDNYVTADSWYRPRDLKNGKWTWASSSDLRPLLSMWDPK 289
 QY 266 ATKAAAYANFWAKEGLISG--SYRONSA--NLDAATQNTQSAIEKKIASGNTWLRDKMSQ 322
 DB 290 QTQIAYLVNMQOGLGTGENYTDSSQESLNLAQTQVVKIETKISQTQQTQWLRLDINS 349
 QY 323 FVKSQNQWSIASENETVYPNDHQGGALLFNSKQTEHANSDWRLNLRNPTFTQTKQKY 382
 DB 350 FYKTOPNWSQTESGAKGKHQGGALLYSNDKTYAVANSYRLNLRPTTSQTKPKY 409

XX AC ADX37279;

XX DT 21-APR-2005 (first entry)

XX DE Streptococcus mutant glucan binding protein B variant #8.

XX DE immunogenicity; immune stimulation; glucan binding protein-B;

KW microparticle; major histocompatibility complex; tooth disease.

XX OS Streptococcus mutans.

XX PN US2005031633-A1.

XX PD 10-FEB-2005.

XX PF 09-MAR-2004; 2004US-00797821.

XX PR 13-APR-1998; 98US-0081550P.

PR 08-JAN-1999; 99US-0115142P.

PR 12-APR-1999; 99US-00290049.

PR 07-MAR-2002; 2002US-0363209P.

PR 08-AUG-2002; 2002US-0402483P.

PR 07-MAR-2003; 2003US-00383930.

XX (SMIT/) SMITH D J.

PA (TAUB/) TAUBMAN M A.

XX Smith DJ, Taubman MA;

XX WPI; 2005-151644/16.

XX New composition comprising a fragment of a glucan binding protein-B (GbpB) that binds to MHC class II protein, and a biocompatible microparticle, useful for producing an antibody (claimed) for immunizing mammals against dental caries.

PT Claim 7; SEQ ID NO 36; 73pp; English.

XX The invention relates to a composition comprising a fragment of a glucan binding protein-B (GbpB) and a biocompatible microparticle, where the fragment binds to a major histocompatibility complex (MHC) class II protein. The composition is useful for producing an antibody for immunizing mammals against dental caries. This sequence corresponds to a Streptococcus mutans GbpB protein of the invention.

XX SQ Sequence 1430 AA;

Query Match 43.9%; Score 3174; DB 9; Length 1430;

Best Local Similarity 47.3%; Pred. No. 5.3e-180;

Matches 677; Conservative 213; Mismatches 436; Indels 104; Gaps 38;

QY 1 MEKNRYKLHKVKKQWVAIGVTT--VTLGFLAGQVVAADTNNDGTSV-----QVKNKV 53

DB 1 METKRYKMKVKKHWTVAVASGLITLTGTLGSSVSAETEQTSDKVVTQKSEDDKAA 60

QY 54 PSDPKFDA----QAQNGQLAQMFKAAQADQATQTSQVSPATDGRVDNVTPAANQPAAN 109

DB 61 SESSQTDAPKTQQAQTEQ-TQAQ-SQANVAD-TSTS----ITKETPSQNIITQANSDDKT 113

QY 110 VANQDVANPATDAGALNRQSAADTSTDGKAVP-----QTSQDPG-----H 149

DB 114 VTNKSEEAQTSERTKQSEEAQTATSSQALTOAKAELTKQRTAAQENKNPVDLAAIEN 173

QY 150 LETVDGKTYVVDANGQRLKXNSVIDGKTYVFDGQTGEAQTDLF----KTQANQDNVDP 205

DB 174 VRQIDGKYIYIGSDGQPKKNFALTANKVLYFDKNTG-ALTDTSQYQFKQGLTKLNN--- 229

QY 206 SYQANNOAVSNEASSPETVDNYLTADSWTRPKILKNGQSQWQASSGDLRPILMTWPD 265

DB 230 DYPHQNQVNFENTSLETIDNYVTADSWTRPKDILKNGKVTATSSSDLRPLMSWPPDK 289

QY 266 ATKAAVANFWAKEGLISG-SYRQNSA--NLDAATQNIQSAIEKKIASSEGNTNWLRDKMSQ 322

DB 290 QTOIAYLNMNOQGLTGENTADSSQESLNLAAGTVQVKIETKISQTOQTOWLRDINS 349

QY 323 FVKSQNWSIASENETVYPNQDHMOGGALLFNSKSDTHANSDWRLNLRNPTFTQKQKY 382

DB 350 FVKTPNWSQTESDTSAGEKHLQCGALLYSNSDKTAYANSDYRLNLRPTPTSQTGPKY 409

QY 383 FTTNYA-GYELLANDVNSNPVQAEQNLHLHYLMNMGDIVMGDKDANFDGVRVDADV 441

DB 410 FEDNSSGGYDFLLANDIDNSNPVQAEQNLHLHYLMNYSIVANDPEANFDGVRVDADV 469

QY 442 VNADLLQIQORDYYKAKYGTQDNEKNAIDHLSLILEAWSGNNDNDYVDKDNFSLSIDNDORS 501

DB 470 VNADLLQIATSDYLKAHYGVDSKSEKNAIHLNLSILEAWSNDNDPOYKDTKGAQLPIDNKLRL 529

QY 502 GMLKAF-----GYASAYRGNLSNLATAGLNKNSA--NPDSDPVNPVYFIRAHSEVOT 552

DB 530 SLLYALTRPLEKDNASKNEIRSGLEPVITNLSLNRSAEGKNSERMANVIFIRAHSEVOT 589

QY 553 RIAKIIREKLKGTNADGLTNLTLDLANKAFDIYNODMNATDKVYYPNNLPMAAYAMLQNK 612

DB 590 VIAKIIKAQI-NPKTDGLT-FTLDELKQAFKLIYNEDMRQAKKCYTQSNIPITAYALMLSNK 647

QY 613 DTVTRVYVYDGYTNGQYMATKTPFYNAIETLLKGRIKYVAGCQ--AVSYKQ-----DW 664

DB 648 DSITRLYYGDMYSDDGQYMATKSPYDAIDTLKKARIKYAAGQDMKITYVEGDKSHMD 707

QY 665 S-SGILTSVRYGKANSASDAGNTRNSGALLNRPFRAYRN--LTLNMGAAHSQ 721

DB 708 DYTGLVTSVRYGTGANEATDQSEATKTQGMVITSNPSSLKLNQNDKVINMGAHKNQ 767

QY 722 AYRPLLSTKGIATYLNDSVDSDROYKYTDSQGNLSFASSELOSVANAQUSGMQVWPV 781

DB 768 EYRPLLSTKGLTSYTDAAKS--LYRKTNDKGLVDFDASDIQGLYLPQVSGYLAVWP 826

QY 782 VGAADNQVRSPTCATKQGNIVHQSALDSOVIYEGFSNFQAFQAPDOYTNVIAKN 841

DB 827 VGASNDQVRSVAASKANATQGVYESSALDSQLIYEGFSNFPDFVKDSDYTNKKAQN 886

QY 842 GDLFKSWGITOPEMAPQVSSBEGTFLDSVLNGYAFSDRYDLAMSKNKYSGKODLANA 901

DB 887 VQLFKSWGTSFEMAPQVSSBEGTFLDSIIQNGYAFEDRYDLAMSKNKYSGKQDMINA 946

QY 902 IKGLQSAQKVLSDLPVNLQNLPEKEVVTATRVNQYQAKSGATINKTPYVANTRSYG- 960

DB 947 VKALHKSQIQVIADWVPQIYNLPKEVVTATRVNDYGEYRSDSEIKNTLYAANTKSNK 1006

QY 961 DYQEOYGGKFLDLDLQKLYPRLSTKQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKYV 1020

DB 1007 DYQAKYGAFJSELAAKYPSIPNRTOISNGKKIDPEKITAOKAKYFNGTNILGRGVY 1066

QY 1021 LSE--GNKYLNIADGKLFPLTVLNNTYQGPQVSANGFISKNGGIHYLDKNGQEVKNRF-K 1077

DB 1067 LKDNASDKYFELKGNQYLPKQMTN-----KEASTGFVNDGNGMTFYSTSGYQAKNSFQ 1121

QY 1078 EISGSWYVYFSDGKMATGKTKIGNDITLFPNGKOLKEGVW--YDGKAYYDDNGRTWT 1135

DB 1122 DAGKNWYFYDNNGHMYYGLQNGEVQYFLSNGVQLRESFLENADGSK-NYFGLGNRYS 1180

QY 1136 NKGFVEFRVGDQKWRVYFNGDGTIAIGLVSLDNRLTYFDAYGYQVKGQTVT-INGKSYTF 1194

DB 1181 N-GYYSF--DNDSKRYFDASGVMAVGLKTINGNTQYFDQDGYQVKGAWITSDGKKRYF 1237

QY 1195 DADQGLVQTDNANPAQQAQWKLKLDGNQWGY-RKDGLLTGEQITDGQKVFQDNGVQ 1253

DB 1238 DDGSGNMAVNRFANDK-----NGDWYILNSDGIALVGQVQTINGKTYFYFGDGRQ 1286

QY 1254 VKGQTATDASGVLYRFDYRDDQHQVKGWYTSDDNNWYNESQVLTGLQTTDGTQVYFD 1313

DB 1287 IKGKIITD-NGKLKYFLANSGLAARNIFATDSQNNWYFGSDGVAVTQSTAGKKLYFA 1345

QY 1314 DKGIOAKGKAVMDENGNLRYFDADSGNMLRDRWK-NVDGNWYFYFNRNGLA 1362

Qy	722	AYRPLHSTKDGATYIYNLSDSVTSRQYKYTDSQCNIUSFSASEIQSVANAQVSGMIQWVP	781
Dd	768	EYRPLLTTTKDGLTSYTSDDAAKS-LYRKTNKDGEFLVDASDIQGVLNPQVSGYLAWVP	826
Qy	782	VGAADNODVRTSPSTQATKOGNIYHQSDALDSOVIEGFNFQAFAGSQDPQYTNAVIAKN	841
Dd	827	VGASDNODVRVAASNKANATGOVYESSALDSQLIEGFNFQDFVTYKSDSYTNKKIAQN	886
Qy	842	GDLFKSWGITQFEMAPQVVSSEDTFLDSVILNGYAFSDRYDLAMSKNNKYGSQDLANA	901
Dd	887	VOLFWSGVTSFEMAPQVVSSESDFSJLDSIIQNGYAFEDRYDLAMSKNNKYGSQDWINA	946
Qy	902	IKGQSAGIKVLSPLVNOLYINLPGEKVWTATRNOYGQAQSGATINKTPYVANTRSYG-	960
Dd	947	VKALHKGGIQVIADWWFPQIYNLPKEKVWTATRNDYGEYRKUSEIKNTLIYAANTXSNKG	1006
Qy	961	DYQBYGCGKFLDDILOKLPLRFSTKQISTCKPIDPSVKITNWSAKYPNGSNILGRGAYV	1020
Dd	1007	DYQAKYCGAFLELAACYPSIFNRQTYSNGKKIDPSEKITAWKAKYPNGTNILGRGVYV	1066
Qy	1021	LSE--GNKYNLANLDKGLPLPTVLNNTYGPQVVSANGFISKNGGIHYLDKNGQEKNRF-K	1077
Dd	1067	LKNWASKYFELKGNQTYLFQMOTN-----KEASTGFVNDGNGTFTYSTGYQAKNSFVQ	1121
Qy	1078	EISGSWTFYFSDGRMATGKTIGINDTYLFPMPNGKQLKEGYW--YDGKAYYYDDNGRTWT	1135
Dd	1122	DAKGNWTFYFDNNGHMVFLOQLANGEVOYFLSNGVOLRESFLENADGSK-NYFGHLGNRYS	1180
Qy	1136	NKGFEFRVDOGDKWRYPFNCGDTTAIGVSLDNRTLFDAYGVGVGTQVTV-INGKSYTF	1194
Dd	1181	N-GYYSF--DNDSKWRYFDASGVNAVGUKTINGNTQYFDQBGYQVKGAWITGSDGKKRYF	1237
Qy	1195	DADQGDIVQTDNANPAPOQAGWKLLGDNQWGY-RKDQOLLTGEBTIDGOKVFFPDNGVQ	1253
Dd	1238	DDGSGNMAVNRFANDK-----NGDWYLYNSDIALVGVTQINGKTYTFYFGDGKQ	1286
Qy	1254	VKGGTATDASGVLFYDRDOGHGVKGWYSTDDNWYVYNESGVLTLGLOTIDGQTYFD	1313
Dd	1287	IKGKIITD-NGKLYTIANSSELARNITFATDSQQNNWTFYFGSDGVAVTGSQTIAGKLIYFA	1345
Qy	1314	DKGIOAKGKVAWDENGNLRYFDADSGNMRLDRWK-NVDGNWNYFNRRNGLA	1362
Dd	1346	SDGKQVKGSGFY-TYNGKHVHYHADBSLGQVNRFEADKDNWYLYDSNGEA	1394

RESULT. 7

RESULI /
AΔ1198042

AAU98042
ID AAU98042 standard: protein: 1430 AA.

ID
XX
AC

DT 27-AUG-2002 (first entry)

DE S. mutans glucosyltransferase GTFD mutant T589E.

XX Glucosyltransferase; GTFD; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amorphous; vacuole; paper manufacture; mutant; mutin.

XX	Streptococcus mutans.
OS	Synthetic.

XX	Key	Location/Qualifiers
FH	Misc-difference	589
FT		

FT
XX
PN US2002031826-A1.

XX
PD 14-MAR-2002.

19-DEC-2000; 2000US-00740274.

XX

1997

07-JUN-1995; 95US-00478704.
07-JUN-1995; 95US-00482711.
07-JUN-1995; 95US-00485243.
16-JAN-1998; 98US-00007999.
16-JAN-1998; 98US-00008172.
20-JAN-1998; 98US-00009620.
11-DEC-1998; 98US-00210361.

(NICH/) NICHOLS S E.

Nichols SE;

WPI; 2002-414332/44.

Glucosyltransferase B or D protein useful for producing a glucan useful as substitutes for and additions to modified starch and latexes in paper manufacture, comprises mutations in specific positions.

Claim 36; Page: 44pp; English.

The invention an isolated protein comprising a glucosyltransferase (GTF) B polypeptide having changes at position from I448V, D457N, D567T, K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/K1014T, I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a GTF D polypeptide having changes at positions from T589D, T589E, N471D, N471D/T589D, and N471D/T589E. Also included are a glucan produced by the GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its complementary polynucleotide, a ribonucleic acid sequence encoding the GTF mutant, an expression cassette comprising the polynucleotide operably linked to a promoter, a vector comprising the expression cassette, host cell introduced with the vector, a transgenic plant comprising the vector, a seed or tuber from the transgenic plant, a paper sizing and/or coating composition comprising a glucan produced in a plant transformed with a gene encoding the mutant GTF, wild type or starch, a latex, thermoplastic molecule or their combinations or glucan and starch where the glucan is produced in the amyloplast and/or vacuole or a maize line deficient in starch biosynthesis, transformed with a gene encoding a glucosyltransferase B or D enzyme, wild-type or mutant and a paper comprising the glucan (paper sizing/coating agent). The vector is useful for producing a glucan in a plant. The method comprises transforming a plant cell with the vector, growing the plant cell under plant growing conditions to produce a regenerated plant and inducing expression of the polynucleotide for a time sufficient to produce the glucan in the regenerated plant, where the vector contains a transit sequence from ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and chlorophyll AB binding protein to produce a transgenic plant, and glucan is produced in the amyloplast of potato or the vacuole of sugar beet. Glucans are useful as substitutes for and additions to modified starch and latexes in paper manufacture. Unlike prior art techniques, which require input materials that produce chemical effluents, paper manufacture utilising the glucan produced by GTF, which utilises biologically produced input materials, is more cost-effective and environmentally friendly. Moreover, glucans also exhibit thermoplastic properties and impart gloss to the paper during coating step. The present sequence represents a GTFD mutant of the invention. Note: The present sequence is not shown in the specification but was created by the indexer using the GTFD sequence appearing as AAU98029 and the information in claim 36.

Sequence 1430 AA;

Query Match 43.8%; Score 3168; DB 5; Length 1430;
Best Local Similarity 47.3%; Pred. No. 1.2e-179;
Matches 676; Conservative 213; Mismatches 437; Indels 104; Gaps 38;
1 MEKNLRYLKHKYKQWVAIGVTT--VTLSFLAGQVVAADTNNDGTSV-----QVKNQW 53

Qy	54	PSDPKFDÄ---	QAQNGQLAQMFKAA	QAQADQAT	TSQVSPAT	DGRVDNQVTPA	QAQPAAN	109
		:	:	:	:	:	:	
Db	61	SESSQTDAPKTKQA	QTEO-TOAQ-SQA	NVAD-TSTS---	ITKETPSONIT	TOANSDDKT	113	

QY 110 VANQDVANPATDAGALNRQSAADTSTDGKAVP-----QTSDPQG-----H 149
DB 114 VNTKSEEAQTSBERTKQSEEAQTASSQALTOAKAELTKQRTQAAQENKNPVDLAAIEN 173
QY 150 LETVDCKTYVVDANGRLKNYSWVIDGKTYYPDQGTGEAQTDL-----KTQANQDNVDP 205
DB 174 VQIDGKYIYGSDGQPKNFALTUVNKKVLYFDKNTG-ALTDTSQYQFKQGLTKLNN--- 229
QY 206 SYOANNOAYSNEASSFETVDNYLTADSWYRPRKILKNGSQWQASSECDLRPIILMTWPPDA 265
DB 230 DYTPHNQIVNFENTSLETIDNYVTADSWYRPRKILKNGKTWTATSSSDLRPLMSWWPDK 289
QY 266 ATKAAAYANFAWEKGLISG-SYRONSA--NLDAAQNIQSAIEKKIASSEGNTNWLKMSQ 322
DB 290 QTQIAYLNTMNQOGLGTGENYTADSSQESLNLAAQTQVQKIETKISQTQQTQWLRLDIINS 349
QY 323 FVKSONOWSIASNETVYPNODHMOGALLFNSKDTEHANSWRLLNRPITFQCKQKY 382
DB 350 FVKTPNWSQTESDTSAGEKHQGLGALLYSNDSKTAYANSYRLNRTPTSTQCKPKY 409
QY 383 FTTNYA-GYELLILANDVDNSNPVQAEQLNHLHLNMGDI VMGDKDANFDGVRVDAVDN 441
DB 410 FEDNSSGGYDFLLANDIDNSNPVQAEQLNHLHLNMGYSIVANDPEANFDGVRVDAVDN 469
QY 442 VNADLLQIORDYKAKYGTQDQNEKNAIDHLSILEAWSGNDNDYVKDNNFSLIDNDQRS 501
DB 470 VNADLLQIASDYLKAHYGVDKSEKNAIHLNLSILEAWSNDNDPOYKDKTGAQLPIDNKLRL 529
QY 502 GMLKAF-----GYASAYRGNLSNLATAGLKNRSA-NPDSDPVNPYFIRAHDSVQT 552
DB 530 SLTYALTREPLEKADSNKNEIRSGLEPVITNSLNRSAEGKNSERMANYIFIRAHDSVQE 589
QY 553 RIAKIIREKLGKTNADGLNTLDDLNKAFDIYNQDMNATDKVYPNNLPMAYAWMLQNK 612
DB 590 VTAKIUKAQI-NPKTDGLT-FTLDELKQAFKIYNEDMRQAKKYTQSNIPYALMLSNK 647
QY 613 DTVTVYVYGDYTDNGQYMATKTPFYNAIETLLKGRIKYVAGQ--AVSYKQ-----DW 664
DB 648 DSITRLYVYGDYSDQYMATKSPYDAIDTLKARIKAYAAQGDQMKITVYEGDKSHMDW 707
QY 665 S-SGILTSVRYGKANSASDAGNTETRNSGMALLNNRPNFRAYRN--LTLNMGAAHKKQ 721
DB 708 DVTGVLTSVRYGTGANEATDQSEATKTQGMAVITSNPSLKLNDKQVIVNMGAAHKNQ 767
QY 722 AYRPLLSTKDGATYLDNSDVDSROYKYVTDQGNLSFASSELQSVANAQVSGMTQVWPV 781
DB 768 EYRPLLSTKDGATYLDNSDVDSROYKYVTDQGNLSFASSELQSVANAQVSGMTQVWPV 826
QY 782 VQAADNQDVRTSPSTQATKQGNIIHQSDALDSQVIYEGFSNFQAFQSPDQYTNVIAKN 841
DB 827 VQASDNQDVRVAASNKANATQVYESSSALDSQVIYEGFSNFQDFTVKDSYDYNKKIAQN 886
QY 842 GDLFKSWGITQFEMAPQVYSSDGTFLDSVIINGVAFSDRYDLAMSKNKYGSKODLANA 901
DB 887 VQLFKSWGITSFEMAPQVYSSDGTFLDSIIQNGVAFEDRYDLAMSKNKYGSKODDMINA 946
QY 902 IKGLSAGIKVLSDLPVNLPGKEVVYATRVNOYGOAKSGATINKTPYVANTRSYG- 960
DB 947 VVALHKSGLQVADWVWDQIYNLPKGEVVYATRVNDYGEYRDKSEIKNTLYAANTKSNK 1006
QY 961 DYQEQYGGKFLDQKLYPRLFSTKQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKTV 1020
DB 1007 DYQAYYGAFLSELAAKYPSIFNRTQISNGKKIDPSEKITAMKAKYFNGTNILGRGVGV 1066
QY 1021 LSE--GNKYLNADGKLFPTVLNNTYQOPQVSANGFTSKNGGHIHVDKNGQEVKNRP-K 1077
DB 1067 LKDNASDKYFELKGNQTLPLPKQMTN-----KEASTGFVNDGNGMTFYSTSGYQAKNSFVQ 1121
QY 1078 ELSGSMWYFSDGKWKATKTKIGNDTYLPMPNGKOLKEGW--YDGKAYYYDDNGRTWT 1135
DB 1122 DAKGNWYIFPDNNGHVMYGLQUNGEVQVPLSLNGFLRESFLENADGSK-NYFCHLGNRYS 1180
QY 1136 NKGFVEFRVDGQDKWRYFNGDGTIAIGLVSLDNRLTYFDAYGYQVKQGTVT-INGKSYTF 1194

DB 1181 N-GYTSF--DNDSKWRFYDASGVMAVGLKTINGTQYFDQDGYQVRGAMITGSDGKRYF 1237
QY 1195 DADQGDVLQVTDNANPAPOGQAGKLLGDNQGY-RKDQQLLTGTEQITDQKQVFFQDNGVQ 1253
DB 1238 DGSGNMVNRFPANDK-----NGDWYILNSDGLALVGQVQTNGKTYTFGQDQKQ 1286
QY 1254 VRKGATDASGVLRFYDRDQGHQVGKMYSTSDNNVYVNESQVLTGLQITIDGQTVYFD 1313
DB 1287 IKGKIITD-NGKLKYFLANSGLARNIFATDSQNNWYFSGDGVAVTGSQTIAGKKLYPA 1345
QY 1314 DKGIQAKGKAVDENGNLRYFADSGNMLDRWK-NVDGNWYFENRGLA 1362
DB 1346 SDGKQVKGSEFV-TYNGKVHYHADSGELOVNRPEADKGNWYILDSNGEA 1394
RESULT 8
AAU98041
ID AAU98041 standard; protein; 1430 AA.
XX
AC AAU98041;
XX
DT 27-AUG-2002 (first entry)
XX
DE S. mutans glucosyltransferase GTFD mutant T589D.
KW Glucosyltransferase; GTFD; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture; mutant; mutein.
XX
OS Streptococcus mutans.
XX
XX Key Location/Qualifiers
FT Misc-difference 589 /note= "Wild-type Thr substituted by Asp"
XX US2002031826-A1.
XX 14-MAR-2002.
XX 19-DEC-2000; 2000US-00740274.
XX 07-JUN-1995; 95US-00478704.
PR 07-JUN-1995; 95US-00482711.
PR 07-JUN-1995; 95US-00485243.
PR 16-JAN-1998; 98US-00007999.
PR 16-JAN-1998; 98US-00008172.
PR 20-JAN-1998; 98US-00009620.
PR 11-DEC-1998; 98US-00210361.
XX (NICH/) NICHOLS S E.
XX Nichols SE;
XX WPI; 2002-414332/44.
DR
PT Glucosyltransferase B or D protein useful for producing a glucan useful
PT as substitutes for and additions to modified starch and latexes in paper
PT manufacture, comprises mutations in specific positions.
XX
PS Claim 36; Page; 44pp; English.
XX
CC The invention an isolated protein comprising a glucosyltransferase (GTF)
CC B polypeptide having changes at position from 1448V, D457N, D567T,
CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,
CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
CC complementary polynucleotide, a ribonucleic acid sequence encoding the
CC GTF mutant, an expression cassette comprising the polynucleotide operably
CC linked to a promoter, a vector comprising the expression cassette, host

OS Streptococcus mutants.
XX Synthetic.

Key Location/Qualifiers

FT Misc-difference 471
FT /note= "Wild-type Aen substituted by Asp"
FT Misc-difference 589
FT /note= "Wild-type Thr substituted by Asp"

PN US2002031826-A1.

XX 14-MAR-2002.

XX 19-DEC-2000; 2000US-00740274.

XX 07-JUN-1995; 95US-00478704.

PR 07-JUN-1995; 95US-00482711.

PR 07-JUN-1995; 95US-00485243.

PR 16-JAN-1998; 98US-00007999.

PR 16-JAN-1998; 98US-00008172.

PR 20-JAN-1998; 98US-00009620.

PR 11-DEC-1998; 98US-00210361.

XX (NICH/) NICHOLS S E.

XX Nichols SE;

XX WPI; 2002-414332/44.

XX Glucosyltransferase B or D protein useful for producing a glucan useful
PT as substituents for and additions to modified starch and latexes in paper
PT manufacture, comprises mutations in specific positions.

XX Claim 36; Page; 44pp; English.

XX The invention an isolated protein comprising a glucosyltransferase (GTF)
CC B polypeptide having changes at position from I48V, D457N, D567T,
CC K104L, D457N/D567T, D457N/D571K, D567T/D571K/K1014T,
CC I448V/D457N/D567T/D571K/K1014T, Y169A/Y170A/Y171A, and K779Q or a
CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
CC complementary polynucleotide, a ribonucleic acid sequence encoding the
CC GTF mutant, an expression cassette comprising the polynucleotide operably
CC linked to a promoter, a vector comprising the expression cassette, host
CC cell introduced with the vector, a transgenic plant comprising the
CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
CC coating composition comprising a glucan produced in a plant transformed
CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
CC thermoplastic molecule or their combinations or glucan and starch where
CC the glucan is produced in the amyloplast and/or vacuole or a maize line
CC deficient in starch biosynthesis, transformed with a gene encoding a
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
CC comprising the glucan (paper sizing/coating agent). The vector is useful
CC for producing a glucan in a plant. The method comprises transforming a
CC plant cell with the vector, growing the plant cell under plant growing
CC conditions to produce a regenerated plant and inducing expression of the
CC polynucleotide for a time sufficient to produce the glucan in the
CC regenerated plant, where the vector contains a transit sequence from
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
CC is produced in the amyloplast of potato or the vacuole of sugar beet.
CC Glucans are useful as substitutes for and additions to modified starch
CC and latexes in paper manufacture. Unlike prior art techniques, which
CC require input materials that produce chemical effluents, paper
CC manufacture utilising the glucan produced by GTF, which utilises
CC biologically produced input materials, is more cost-effective and
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
CC properties and impart gloss to the paper during coating step. The present
CC sequence represents a GTFD mutant of the invention. Note: The present
CC sequence is not shown in the specification but was created by the indexer
CC using the GTFD sequence appearing as AAU98029 and the information in
CC claim 36

XX	SQ	Sequence 1430 AA;
		Query Match 43.7%; Score 3163; DB 5; Length 1430;
		Best Local Similarity 47.2%; Pred. No. 2.4e-179;
		Matches 675; Conservative 214; Mismatches 437; Indels 104; Gaps 38;
QY	1	MEKNRYKHLHKVKKQWVAIGVTT--VTLSFLAGGVVAAADTNNDGTSV-----QVNMV 53
DB	1	METKRRYKHKVKKHWTVAVASGLITLTGTTTLLGSSVSAETEQTQSKVVTQKSEDDKAA 60
QY	54	PSDPKFDA-----QAQNGQLAAMFKAAQADQATQTSQVSPATGRVDNQVTPPAANQPAAN 109
DB	61	SESSQTDAPKTKQAQTEQ--TQAQ--SQANVAD--TSTS-----ITKETPSQNTITQANSDDKT 113
QY	110	VANQDVANPATDAGALNRQSAADTSTDGKAVP-----QTSQDPG-----H 149
DB	114	VNTKSEEAQTSEERTKQSEEAQTASSQALTOAKAELTKQRTAAQENKNPVDLAAIPN 173
QY	150	LETVDGKTYTVDANGQRLKNYSMVIDGKTYTDFDQTGEAQTDLF----KTGOANQDNVDP 205
DB	174	VKQIDGKYVYIGSDGQPKKNFALTVNKVLVFDKNTG--ALTDTSQYQFKQGLTKLNN--- 229
QY	206	STQANNQAYNEASPFETVDNYLTADSWTRPRKILKNQSGSQWASSEGDLRPLMTWMPDA 265
DB	230	DYTPHNQIVNFENTSLETIDNYVTADSWTRPRKILKNQSGSQWASSEGDLRPLMTWMPDK 289
QY	266	ATKAAVANFWAKEGLISG--SYRQNSA--NLDAAQNIQSAIEKKIASSEGNWLRDKMSQ 322
DB	290	QTOIATLNTMNOOGLTGENTYADSSQESLNLAQTQVKIETKISQTOQTQWLRLDIINS 349
QY	323	FVKSQNQSIASENETVYPNQDHMQGALLFNSKDETHANSWRLLNRPFTQTGKQKY 382
DB	350	FVKTPQMNWSQTESDTSAGEKHLQGGALLYNSDKTAYANSYRLNRTPTTSQTGPKY 409
QY	383	FTTNVA--GYELLANDVNSNPVQABQLNHLHYLNMGDIIVMGDKANPDGVRVDAVDN 441
DB	410	FEDNSGGYDFLLANDIDNSNPVQABQLNHLHYLNMGDIIVMGDKANPDGVRVDAVDN 469
QY	442	VNADLLOIORDYKAKYGTQDNEKNAIDHLSILEAWSGNDNVVQDNFSLSDNDORS 501
DB	470	VADLLQIATSDYLKAIHYGVKDKSEKNAINHLSILEAWSNDPQYNKDTKAQFLDKRL 529
QY	502	GMLKAF-----GYASAYRGNLSNLATAGLKNRSA--NPDSPVPVNYVYFIRAHDEVQT 552
DB	530	SLLYALTRPLEKADSKNKEIRSGLEPVITNSLNRSAGEKNSERMANYIFIRAHDEVQD 589
QY	553	RTAKIIREKLGKTNADGLTNLTLDDLKAFDIYNQDMNATDKVYYPNNLPMAVWMLQNK 612
DB	590	VIAKIITKAQI--NPKTDGLT--FTLDELKQAFKIYNEQDMRQAKKKTQSNIPYALVLSNK 647
QY	613	DTVTVVYGDVYTDNGOYMATKTPFYNAETILLKGRIVYVAGGQ--AVSYKQ-----DW 664
DB	648	DSITRLYGDVYDSDGQYMATKSPYDAIDTLKARIKVAAGGQDMKITYVEGDKSHMDW 707
QY	665	S--SGILTSVRYGKGANASDAGNTETNRSGMALLNNRPNFRAYRN--LTLNMGAAHKSQ 721
DB	708	DYTGVLTSVRYTGANEATDQSEATKTQGMAVITSNPFLKLNQNDKVIWVMGAHKNQ 767
QY	722	AYRPLLLSTKGIATYLANDSDVDSDRQYKYVYTDQSGNLSFSAEQLQSVANQVSGMIQVWP 781
DB	768	EYRPLLLATKDLGTSYSDAAAKS--LYRKTNDKGLVFDASDIQGYLNPQVSGYLAVWP 826
QY	782	VCAANDQVTRSPSTQATKQDGNIIHQSDALDSQVLYEGFSNFQAQSPQDTNVAIQN 841
DB	827	VGASNDQVTRVAASNKANATQGVYESSGALDSQLIYEGFSNFQDFVTKDSDTNNKIAQN 886
QY	842	GLFKSWGITQEPMAPOVYSSSDGTFLDSVLNGVAFSDRYDLAMSKNNKYSKODLANA 901
DB	887	VOLFSGWGTSEMAPQOVYSSSDGTFLSIIQNGVAFEDRYDLAMSKNNKYSQQQMIINA 946
QY	902	IKGLQASAGIKVLSDLVPNQNLNLPKGKVVVTAATRVNQYQAKSGATINKTPPYVANTRSYG- 960

Db 947 VKALHSGIOVIADWVPDQIYNLPGKEVVTATRVNDYGEYKDSBKNTLYAANTKSNCK 1006
Qy 961 DYQEQYGGKFLDLOKLYPRLFTSTKQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKYV 1020
Db 1007 DYQAKYGGAFSLAELAAKYPFIENRTQISNGKKIDPSEKITAMKAKYFNGTNILGRGVGV 1066
Qy 1021 LSE--GNKYLNLADGKLFPLTPTVNLNTYQOPQVSANGFISNGGIHYLDKNGQEVKRP-K 1077
Db 1067 LKDNASDKYFELKGNQTYLPKQMTN-----KEASTGFVNDGNGMTFYSTSGYQAKNSFPVQ 1121
Qy 1078 EISGSMYYFSDGKMATGKTKLGNDRYLFMPNGKQLKEGVW--YDGKKAYYYDDNGRWTW 1135
Db 1122 DAKGNWYFDNNGHMYVGLQUNGEVOYFSLNSGVQURESFLNADGSK-NYFCHLGNRY 1180
Qy 1136 NKGFEVFRVGDGKRWYFNGDGTIAIGLVSLNRLYFPDAYGYQVKGQTVT-INGKSYTF 1194
Db 1181 N-GYYSF--DNDSKRWYFASGVMAVGLKTINGNTQYFPDQGVQVKGAWITGSDGKRYF 1237
Qy 1195 DADQGLVOTDNANPAPOQAGCKLLGDNQWGY-RKDGQLLTGEBQITDQKVPFQDNGVQ 1253
Db 1238 DDGSGNMAVNFANDK-----NGDWYLLNSDGIALVGVTQINGKTYFFGQDGKQ 1286
Qy 1254 VKGTTATDASGLVRFVDRDQGHQVGKGYSTSDDNWYVNESQVLTGTQITIDGOTVYFD 1313
Db 1287 IKGKIITD-NGKLKYPFLANSGLARNIFATDSQNNWYFPGSDGVAVTGSQTTAGKRLYPA 1345
Qy 1314 DRGIOAKGKAVMDENGLRYFDADSGNMLRDRWK-NVDGNWYTFNRRGLA 1362
Db 1346 SDGKQVKGFSV-TYNGKVHYHADSGELQVNRFEADKQGNWYLLDSNGEA 1394

RESULT 10
AAU98045
ID AAU98045 standard; protein; 1430 AA.
XX
AC AAU98045;
XX
DT 27-AUG-2002 (first entry)
XX
DE S. mutans glucosyltransferase GTFD mutant N471D/T589E.
XX
KW Glucosyltransferase; GTFD; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture; mutant; mutein.
XX
OS Streptococcus mutans.
OS Synthetic.

XX Key Location/Qualifiers
FH Misc-difference 471 /note= "Wild-type Asn substituted by Asp"
FT Misc-difference 589 /note= "Wild-type Thr substituted by Glu"
FT
XX US2002031826-A1.
XX
XX 14-MAR-2002.
XX
XX 19-DEC-2000; 2000US-00740274.
XX
XX 07-JUN-1995; 95US-00478704.
XX 07-JUN-1995; 95US-00482711.
XX 07-JUN-1995; 95US-00485243.
XX 16-JAN-1998; 98US-00007999.
XX 16-JAN-1998; 98US-00008172.
XX 20-JAN-1998; 98US-00009620.
XX 11-DEC-1998; 98US-00210361.

XX (NICH/) NICHOLS S E.
XX
XX Nichols SE;
XX
XX WPI; 2002-414332/44.
DR

XX
PT
PT
PT
XX
PS
Claim 36; Page; 44pp; English.
XX

The invention an isolated protein comprising a glucosyltransferase (GTF) B polypeptide having changes at position from 1448V, D457N, D567T, K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/D1014T, D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a GTF D polypeptide having changes at positions from T589D, T589E, N471D, N471D/T589D, and N471D/T589E. Also included are a glucan produced by the GTF mutant, an isolated polynucleotide which encodes F1 or F2, or its complementary polynucleotide, a ribonucleic acid sequence encoding the GTF mutant, an expression cassette comprising the polynucleotide operably linked to a promoter, a vector comprising the expression cassette, host cell introduced with the vector, a transgenic plant comprising the vector, a seed or tuber from the transgenic plant, a paper sizing and/or coating composition comprising a glucan produced in a plant transformed with a gene encoding the mutant GTF, wild type or, starch, a latex, thermoplastic molecule or their combinations or glucan and starch where the glucan is produced in the amyloplast and/or vacuole or a maize line deficient in starch biosynthesis, transformed with a gene encoding a glucosyltransferase B or D enzyme, wild-type or mutant and a paper comprising the glucan (paper sizing/coating agent). The vector is useful for producing a glucan in a plant. The method comprises transforming a plant cell with the vector, growing the plant cell under plant growing conditions to produce a regenerated plant and inducing expression of the polynucleotide for a time sufficient to produce the glucan in the regenerated plant, where the vector contains a transit sequence from ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and chlorophyll AB binding protein to produce a transgenic plant, and glucan is produced in the amyloplast of potato or the vacuole of sugar beet. Glucans are useful as substitutes for and additions to modified starch and latexes in paper manufacture. Unlike prior art techniques, which require input materials that produce chemical effluents, paper manufacture utilises the glucan produced by GTF, which utilises biologically produced input materials, is more cost-effective and environmentally friendly. Moreover, glucans also exhibit thermoplastic properties and impart gloss to the paper during coating step. The present sequence represents a GTFD mutant of the invention. Note: The present sequence is not shown in the specification but was created by the indexer using the GTFD sequence appearing as AAU98029 and the information in claim 36

XX
SQ Sequence 1430 AA;

Query Match 43.7%; Score 3163; DB 5; Length 1430;
Best Local Similarity 47.2%; Pred. No. 2.4e-179;
Matches 675; Conservative 214; Mismatches 437; Indels 104; Gaps 38;

Qy 1 MEKNLRYKLHKVKQWVAIGVT--VTLFLAGGVVAAADTNNDGTSV----QVKNKV 53
Db 1 METKRYKMKHKVKKHVTVAVASGLITLTGTLGSSVSAETEQTSKDVTKQKSEDDKAA 60
Qy 54 PSDPKFDA---QAQNGQLAQMFKAAQADQATATTSQVSPATDGRVDNVTPAANQPAAN 109
Db 61 SESSQTDPKTKQAQTEQ-TQAQ-SQANVAD-TSTTS----ITKETPSQNIITQANSDDKT 113
Qy 110 VANQDVANPATDAGALNRQSAADTSTDGKAVP-----QTSDPG-----H 149
Db 114 VTNTKSEEAQTSSEERTKQSEEAQTASSQALTKQAKELTKQRTAAQENKNPVDLAAIPN 173
Qy 150 LETVDGKTYVDANGORLKNYSMWIDGKTYYPDGTGEAQTDLP----KTGAQANQVND 205
Db 174 VKQIDGKYIYIGSDGQPKGNFALTVNNKVLVYDKNTG-ALTDTSQYQFKQGLTKLNN--- 229
Qy 206 SYQANNQAYSNEASSFETVDNYLTADSWYRPRKILKNGQSQWASSSGDLRPLMTWPPDA 265
Db 230 DYTPHNQIVNFENTSLTIDNTVTDADSWYRPRKILKNGKGTATSSSEDLRPLMSWPPDK 289
Qy 266 ATKAAYANFWAKEGLISG-SYQNSA--NLDATQNIQSAIEKKIASEGNTWLRDKMSQ 322

Db 290 QTOIAVLTWNOQGLGTGENYATDSSOESLNLAAQTQVVKIETKISQTOOTQRLRDIINS 349
QY 323 FVKSQNWSIASNETVYQNDHMOGALLFSNSKDTTEHANSDWRLNLRNPTFQTGCKQY 382
Db 350 FVKQTPNWSQTESDTSAGEKHLQGGALLYSNSDKTAYANSYRLNLRNPTFSQTKPKY 409
QY 383 FTTNYA-GYELLANDVNSNPVQAEQLNHLHLYLNMWGDIVMGDKDANFDGVRVDAVDN 441
Db 410 FEDNSSGGYDFLLANDIDNSNPVQAEQLNHLHLYLNMWGDIVMGDKDANFDGVRVDAVDN 469
QY 442 VNADLLQIQORDYYKAYGTDONEKNAIDHLSILEAWSGNDNDVYKDNFSLSIDNDQRS 501
Db 470 VDADLLQIASDYLUKAYGVDSKSEKNAINHLSILEAWSNDNDPQYNKDTKAQLPIDNKLRL 529
QY 502 GMLKAF-----GYASAYRGNLSNLATAGLNKRS-A-NPDSDPVPVNYVPIRAHDESVOT 552
Db 530 SLLYALTRPLEKADSKNEIRSGLEPVITNSLNRSAEGKSERMANYIPIRAHSEVOB 589
QY 553 RTAKIIREKLGKTNADGLTNLDDLNKAFDINQDMNATDKVYPNNLPMAYMWLQNK 612
Db 590 VIAKIIKAQI-NPKTDGLT-FTLDELKQAFKIYNEDMRQAKKKYQTSNIPATAVALMSNK 647
QY 613 DTVTRVYGYDMYTDNGQYMATKTPFYNAIETLLKRIKIVAGGQ--AVSYKQ-----DW 564
Db 648 DSITRLYGYDMYDSDQYMATKSPYYDAIDTLKARIKAYAAGQDMKITYVEGDKSHMDW 707
QY 665 S-SGILITSVRYGKANSASDAGNTESTRNSGMALLNNRPNFRAYRN--LTLNNGAAHKSQ 721
Db 708 DYTGLVTSVRYGTGANEATDQSEATKTQGMVITSNFSCLKNQNDKVINMGAAHKNQ 767
QY 722 AYRPLLSTKOGIATYLNDSVDSDQYKYTDSQGNLSPSASELSQVANAQVSGMIOVWYP 781
Db 768 EYRPLLSTKOGIATYLNDSVDSDQYKYTDSQGNLSPSASELSQVANAQVSGMIOVWYP 826
QY 782 VCAADNODVTSPTSQATKQGNHLYHQSADLDSQVIEGFSNFQAFQSPDQYTNVIAKN 841
Db 827 VGASNDQDVRVAASKANATQGVYESSSALDSQLIYEGFSNFQDFTVKDSYTNKKIAQN 886
QY 842 GDLFKSWGITQEPMAPOVYSSSDGTFLDSVILNGVAFSDRYDLAMSKNNKYSKODLANA 901
Db 887 VOLFKSWGITSFEMAPOVYSSSDGTFLDSVILNGVAFSDRYDLAMSKNNKYSKODLANA 946
QY 902 IKGLQAGIKVLSDILVFNQYLNLPKGVVTVATRVNQQYQAKSGATINKTPYVANTRSYG- 960
Db 947 VKALHKSGLQIADWVPDQIYNLPKGVVTVATRVNQQYQAKSGATINKTPYVANTRSYG- 1006
QY 961 DYQEQYGGKFLDQKLPRLPSTQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKYV 1020
Db 1007 DYQAKYGGAFSLAELAAKYPISFNRTQISNGKKIDPSEKITAWKAKYFNGTNILGRGVY 1066
QY 1021 LSE--GNKYLNLADGKFLPLTVLNTYQPOVSANGFISKNGIHYLDKNGQEVKNRF-K 1077
Db 1067 LKDNASDKYFELKGNQTLPLPKQMTN-----KEASTGFVNDGNGMTFYSTGYQAKNSFVQ 1121
QY 1078 EISGWYTFDSQKMATKTKICNDTYLPMNGKQLKEGW--YDGKXAYYDDNGRTWT 1135
Db 1122 DAKGNWYTFDNGHVMYGLQNGEVQYFLSNGVQLRESFLENADGSK-NYEGHLGNYS 1180
QY 1136 NKGFEVFRVGDQKWRYNFGDGTIAIGLVSLNRLTYFDAYGYQVKGQTVT-INGKSYTF 1194
Db 1181 N-GYYSF--DNDSKWRVFDASGMAVGLKTINGNTQYFDQGYQVKGAWITGSDGKRYF 1237
QY 1195 DADQGLVQTDNANPAPQOAGKLLGDNGWY-RKDGOLLTGEOTIDGQKVPFODNGVQ 1253
Db 1238 DSGSNMAMVNFANDK-----NGDWYTLNSDGIALGVQTINGTKTYFYQDGKQ 1286
QY 1254 VKGFTATDASGLVRFYRDQGHQVKGWYSTDDNNWYVNESGQVLTGLQITIDGQTVYFD 1313
Db 1287 IKGKIITD-NGKLYFLANSGLARNIFATDSQNNWYFGSDVAVTGSQTIAGKLIYFA 1345
QY 1314 DKGIOAKGKAVMDENGLRYFDADSGNMLRDRWK-NVDGNWYFNRNGLA 1362

Db 1346 SDGKQVKGFSV-TYNGKVHYHADSGELQVNRFEADKGNWYVYLDNNGEA 1394
RESULT 11
ADD93658
ID ADD93658 standard; protein; 1554 AA.
XX ADD93658;
XX AC
XX 29-JAN-2004 (first entry)
XX Streptococcus sobrinus glucosyltransferase-U.
XX Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.
XX Streptococcus sobrinus.
XX WO2003075845-A2.
XX 18-SEP-2003.
XX 07-MAR-2003; 2003WO-US006962.
XX 07-MAR-2002; 2002US-0363209P.
XX 08-AUG-2002; 2002US-0402483P.
XX (FORS-) FORSYTH INST.
XX Smith DJ, Taubman MA;
XX WPI; 2003-845091/78.
XX Composition useful as vaccines for dental caries comprises a fragment of
PT a glucan binding protein-B binding to a major histocompatibility complex
PT class II protein.
XX Claim 16; Page 15; 49pp; English.
XX The present sequence is the protein sequence of Streptococcus sobrinus
CC glucosyltransferase-U (GTF-U). Peptide fragments of GTF-U, especially
CC from the catalytic domain of the polypeptide, can be used in immunogenic
CC compositions and subunit vaccines for dental caries. These compositions
CC comprise a major histocompatibility complex (MHC) class II protein-
CC binding peptide from S. mutans glucan binding protein-B (GbpB)
CC covalently linked with a peptide fragment of a streptococcal
CC glucosyltransferase. The compositions are used in a claimed method of
CC eliciting production of an antibody in a mammal. Diepitopic or
CC multipitopic polypeptides can be prepared synthetically or by
CC recombinant DNA technology. Antibodies raised against MHC class II
CC binding fragments of GbpB can be used in passive immunisation.
XX Sequence 1554 AA;
SQ

Query Match 42.6%; Score 3080.5; DB 7; Length 1554;
Best Local Similarity 43.9%; Pred. No. 2.3e-174;
Matches 667; Conservative 213; Mismatches 446; Indels 193; Gaps 34;

QY 1 MEKRLRYKLHKVKKQWVAIGVTTVTLSPFLAGQVAA-----DTNNDG 44
Db 1 MEKRLRYKLHKVKKHMTTIAVASIGLVSLVGAGTVSAEDKVANDTTAQATGVDTGQDQA 60
QY 45 TSVQVKNVPSDPKFDQAQNGQLAQMFKAAQADQ---TATSOVSPATGVRVDNQVT 100
Db 61 TTNDANTTTTDTTADQASANTNQ-DQAGSDQSNNDQAKQDTANTDRNADNSQTDN---- 116
QY 101 PAANQPAANVANQDVANPATDAGALNRQSAADTST-----DGKAVPQTSDPG-----HLE 151
Db 117 -----NQATDQATSPATDGTSTVQRDAANVATAADQEGQTASEQEKSAALSILDNVY 168
QY 152 TVDGKTYVDANGQRLKYSWVIDKTYTFDQGTGEAQTDLKPTQANQDNVDPDSYQANN 211
Db 169 LIDGKYTYVQADGSYKKNFAITVNGQMLYFSDTDGALSSTSYSPSQGTTLNLDVDFSSH 228

Db 1 MEKKLHYLKHVKKHVVTTAVASIGLVSLVAGACTVSAEDKXVANDTTAQTAVGTGVDGQDQA 60
QY 45 TSVQVNMVSPSPKFDQAQNGQLAQAMFKAANQADQ-----TATSOVSPATGRVDNQVT 100
Db 61 TTNDANTNTTDTADQASANTNQ--DQAGSDQSNQDQAKQDPTANDRNQADNSQTDN--- 116
QY 101 PAANQPAANVQDVANPATDAGALNRQAADTST-----DGKAVPQTSQPG-----HLE 151
Db 117 -----NOATDQATSPATDGTSTVQRDAANVATAADQEGQTAPSEKSAALSIDNVK 168
QY 152 TVDGKTYVYVDANGORLKNYSWIDGKTYVFDGQTEAQTDLPKTGOANQDNVPSQANN 211
Db 169 LIDGKYVYVQAGSGYKKNFAITVNGMLYFDSDTGALSSTYSFGTTLNVDFSSH 228
QY 212 QAYSNEASSFETVDNYLTADSYRRKILKNGSQWASSEGDLRPLMTWPDAAATKAAY 271
Db 229 KAYDSTAKSFELVNGVLTANSWYRPAIILRNQGTWEASNENLRLPVLMSWEPDKDTQVAY 288
QY 272 ANFW-+-AKGLISGYSRONSANLDAATONIQIAIEKKIASEGTNWLRDNKQSFVKS 326
Db 289 VNYMKNYLSANETEVNTNETSQVDLKEA--QSIQTKIEQKITSDNSTQWLRTAMEAFVAA 346
QY 327 QNOWSIASENETVYNQDHHMOGALLFNSKDEHANSRWLRNENPTFTGKQKYFTT- 385
Db 347 QPKWNNSTEN---FNKGDHLGGALLYTNSDLTPWANSDYRLNLTPTQQDGTKKYFTG 403
QY 386 NYAGYELLANDVNSNPVQAEQLNLHLYLNNWGDIVMGDKDANFDGVRVDAVDNVDNAD 445
Db 404 GEGGYEFLLSNDVNSNPVQAEQLNLHLYLNNWGDIVMGDKDANFDGVRVDAVDNVDNAD 463
QY 446 LLIQORDYYKAKYGTQDQEKNAIDHLSILEAWSGNDNDYVQDQNNFSLSDNDQSRGMLK 505
Db 464 LLIQVSYNFKYKQVYTDSEANALAHISILEAWSLNDQYNEGTNGTALSIDNSSRLTSLA 523
QY 506 AFGYASAVRGNLSNLATAGLKNRSNPDS--DPVNYVPIRAHDSVQTRIARIKELG 563
Db 524 VLTKPGQRIIDLNLISSEVKNERANDTAYGDTIPTYSFVRAHDSVQTRIARIKELI- 582
QY 564 KTNADGLNLTLDLNLKAFDIYNQDMNATDKVYYPNNIPMAVAMLQNKDVTTRYVYGD 623
Db 583 DTNSDGYT-FTLQDLKDAFKIYEDMAKVNKTYTHYNIPAYALLSNWESVPRVYGD 641
QY 624 YTDNGQYMATKTPFNATIELLKGRIKYVAGQAVS-YKQDMSGILTSVRYGKGANSAS 682
Db 642 YTDGQYMAKSPYDAIATMLQGRIAVYSGQSEEVHKVNGNQLSSVRYGQDLMSAD 701
QY 683 DAGNTE--TNSGSMALLINRPNF-RAYRNLTLNMGAAHKSQAYRPLLSTKDGATYND 740
Db 702 DTQGTDLSTSLVTLVNDPNDLIDGGDSLTVNMGRAHQAYRPLILGTGQVQSLKD 761
QY 741 SDVSRQYKTYDSQNLSPSASELOSVANAQVSGMIQVWPVGAADNODVRTSPSTQATK 800
Db 762 SDTHI--VKYTDANGNLFTADDIKGYSTVDMSGYLAVVWPVGAKDQDVRVAADTNOKA 819
QY 801 DGNVHQSDALDSQVYIEGFSNFQAFQSPDQYTNNAVIAKNGDLFPKSGITQFEMAPQV 860
Db 820 DGKSLKTSAAALDSQVYIEGFSNFQAFQSPDQYTNNAVIAKNGDLFPKSGITQFEMAPQV 879
QY 861 SSEDGTFDLDSVLNGLAYAFSDRYDLAMSKNKGYSQDLANAIKGLQSAIGIKVLSLDVNO 920
Db 880 SATDGSFLDSTIQNGYAFSDRYDLAMSKNKGYSQDLANAIKGLQSAIGIKVLSLDVNO 939
QY 921 LYNLPGEVVTATRVNQYQAKSGATINKTPVANTRSYG-DYOSQYGGKFLDLDLQKLYP 979
Db 940 IYQLPGEVVTATRVNQYQAKSGATINKTPVANTRSYG-DYOSQYGGKFLDLDLQKLYP 999
QY 980 RLFTSTQIKGPIKIDPSVKITNWSAKYFNGSNILGRAGYVLSG--NKYLNLDADKGLFL 1037
Db 1000 DMFTVMNISTKPIKIDPSVKITNWSAKYFNGSNILGRAGYVLSG--NKYLNLDADKGLFL 1059
QY 1038 PTVLNTNYGQVQVANSANGFISKNGGHIHLDKNGQEVKNRPFKEISGWSYFYDSGKQATG-- 1095
Db 1060 PA-----STGQDNQAKTGYFDGTGMAYYSTSGNAVNSFIVEGGHYHYFYFDKDHMTGYS 1115

QY 1096 KTKIGNDTYLFMPNGKQLKEGVVYDGK-KAYYYDDNGRWTWNG-----FVFRVVDGQDK 1149
Db 1116 KAEDGND-YFPLNGIQMRDAIYQDAQNSYYGRTGILY--KGDWYFVDPNNANKTV 1172
QY 1150 WRYFNGDGTIAIGLVSLNDRNLTLYFAYGYQVKGQVTTNGSKSYTFDADQG----- 1199
Db 1173 FRYFANNVMAIGYRNMYGQTYFYFDENGFQAKQLLTDKGTHTYFDEDNAMAKNKFVNV 1232
QY 1200 --DLVQTNANPAPOG----- 1214
Db 1233 GDDWTYDMONGNAVAGQYFVNNQIILYFNPETGVQVKGQFITDAQGRTSYDANSALKSS 1292
QY 1215 -----AGWKLLGDN--QWGY--RKDQQLLTGEQTIDGQKVFQ-DN 1250
Db 1293 GFPTENGSDWYVAENGYVYKGFQVAENQDQWYFDTTGKQAKGAQVDRDLFNPDS 1352
QY 1251 GVQYKGGTATDASGVLRYFDRDQGHQVKGWYSTSDNNWYVNESGVULTGLQTD----- 1306
Db 1353 GVQYKGGTATDASGVLRYFDRDQGHQVKGWYSTSDNNWYVNESGVULTGLQTD----- 1412
QY 1307 -----GQTVYFD--DKGLOAKGKAVMDENGNLRYFDADSGNMLRDRWKN 1348
Db 1413 HFDEVTGQQAAGAAALVNGQQLYFVDVSGIQVKGDFVTDGQNTSYDVNSGD-----KK 1466
QY 1349 VDG-----NWYFNRNG 1360
Db 1467 VNGFPTGDNWYVYADGOG 1485

RESULT 13

ADD93657
ID ADD93657 standard; protein; 1590 AA.
XX AC ADD93657;
XX DT 29-JAN-2004 (first entry)
XX DE Streptococcus sobrinus glucosyltransferase-I.
XX KW Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.
XX OS Streptococcus sobrinus.
XX PN WO2003075845-A2.
XX PD 18-SEP-2003.
XX PF 07-MAR-2003; 2003WO-US006962.
XX PR 07-MAR-2002; 2002US-0363209P.
XX PS 08-AUG-2002; 2002US-0402483P.
XX PA (FORS-) FORSYTH INST.
XX PI Smith DJ, Taubman MA;
XX PS WPI; 2003-845091/78.
XX
XX Composition useful as vaccines for dental caries comprises a fragment of a glucan binding protein-B binding to a major histocompatibility complex class II protein.
XX Claim 16; Page 14; 49pp; English.

CC The present sequence is the protein sequence of Streptococcus sobrinus glucosyltransferase-I (GTF-I). Peptide fragments of GTF-I, especially from the catalytic domain of the polypeptide, can be used in immunogenic compositions and subunit vaccines for dental caries. These compositions comprise a major histocompatibility complex (MHC) class II protein-binding peptide from S. mutants glucan binding protein-B (GbpB) covalently linked with a peptide fragment of a streptococcal glucosyltransferase. The compositions are used in a claimed method of

CC eliciting production of an antibody in a mammal. Diepitopic or
CC multiepitopic polypeptides can be prepared synthetically or by
CC recombinant DNA technology. Antibodies raised against MHC class II
CC binding fragments of GbpB can be used in passive immunisation.
XX
SQ Sequence 1590 AA;

Query Match 41.9%; Score 3031; DB 7; Length 1590;
Best Local Similarity 41.6%; Pred. No. 2.1e-17;
Matches 672; Conservative 217; Mismatches 436; Indels 292; Gaps 36;

QY 1 MEKNRYKLHKVKKQVAGVITVTN-LGFLAGQVVAADTNNDGTSVQVKNKVPSPDKF 59
DB 1 MEKNRYKFKHKKVKKVLSVASATMLASALGASVASADTDITASDDSNQA--VVTGD--- 55

QY 60 DAQAQNGQLAQAAMFAAKQAQDQATATQSPATDGRVDNQVTPAANQPAANVANQDV---- 115
DB 56 --QTNNQATDQTSIAA-----TATSEQSASTDAATDQ--ASAAEQTQGTASTDTAAQT 106

QY 116 ---ANPA-----TDAGALNRQSAADTSTGKAVPQTSQPOGHLETVDGKTYIV 160
DB 107 TTNAEAKVPTPENENQGTDEMLAEAKNVATAESD--SIPDLAKMSNVKQVDGKYIY 164

QY 161 DANGQRLKNYSVIDGKTYVFDGQ--EAQTDLPKTQGANQDNVPSDYAANQAYSN 216
DB 165 DQDGNVKNFAVSGDKIYIFD-ETGAYKDTSKVDADKSSASVQNA-TIFAANNRAYST 222

QY 217 EASSPETVDNYLTADSWRPRKILKNGOSQWASSEGDLRPLMTWMPDAATAKAAVANFWA 276
DB 223 SAKNEAVDNYLTADSWRPRKILKNGOSQWASSEGDLRPLMTWMPDAATAKAAVANFWA 282

QY 277 KEGLSGSY--RONSANLDAATQNTQSAIEKKIASEGNTNMLRDKOSQFVKSQONWSTAS 334
DB 283 KVGIGDKTYTAETSQADLTAAAEVQARIEQKITSENNTKMLREAISAFVKTQPOWNGES 342

QY 335 ENETVYQNDHMOGALLFNSKQD--TEHANSWRLLNBNPTFOTCK-QKYFTTN----YA 388
DB 343 EK----PYDDHLQNGALLFDNQTDTPTQSNRYLLNLTPTNTQTSLSRSRFTYNDPLG 398

QY 389 GYELLANDVNSNPVQAEQLHLHYLMNMWDGIVMGDKDANFDGVRVDADVNNADLLQ 448
DB 399 GYDFLLANDVNSNPVQAEQLHLHYLMNMWDGIVMGDKDANFDGVRVDADVNNADLLQ 458

QY 449 IORDYKAKYGTQNEKNAIDHLSILEAWGNDNDYVKDQNNFSLISINDQBSGMLKAPG 508
DB 459 ISSDYLKAAYGIDKKNKANNHVSIVEASDNDTPYLDHDDGNLNMNDNKFRLSLWSLA 518

QY 509 YASAYRGNLNLATAGLKNRSANP--DSDPVNPNVYFIRAHDSQVOTRIAKIREKLGKTN 567
DB 519 KPLDKRSGNLNLIHNSLVDRVDDREVETVPSYSPARAHDSQVOTRIAKIREKLGKTN 577

QY 568 DGLTNLTDLDDLNKAFDIYNQDNNATDKVYPPNPLMAYAMQLQNDKTVTRVYGDYMTDN 627
DB 578 FGYSFTQEEIQAFKIYNEDLKKTDKXTHYNVPLSYLLTNKSGIPRVYVYGMFTDD 636

QY 628 GQYMATKTFYMAIFLTKGRIVYAGQVSVYKQDWSGILTSVRYKGSANASDAGNT 687
DB 637 GQYMAKNTVYDAIESLLKARMKYVSGGQAMQNYQIGNGEILTSVRYKGSALKQSDKGA 696

QY 688 ETRNSGMALLINRNPNFRA-YRNLTLNMGAAHKSQAYRPLLLSTKDGATYLNDSQDVS 745
DB 697 TTRTSGVGVNMGQNFSLDGVKVALNMGAAHANQYERALMVSTKDGATYLNDSQDVS 756

QY 746 ROYKYTDSQGNLSFSAELQSVANAQVSGMIQVWVPVGAADNQDVRTSPSTQATKDNIT 805
DB 757 GLVKRTDENGILYFLNDDLKGVANQVSGFLQVWVPVGAADNQDVRTSPSTQATKDNIT 816

QY 806 HQSDALDSQVIYEGSNNFOAFQSPDYTNNAVIAKNGDLFKSWGITQFEMAPQYVSSBDG 865
DB 817 HODAAMDNRVMEPEGSNNFOAFQSPDYTNNAVIAKNGDLFKSWGITQFEMAPQYVSSBDG 876

QY 866 TPLDSVILNGYAFSDRYDLAMSKNNKYSQDLANAIKGLQSAKIKVLSLDVLPNOLYNLP 925
DB 866 TPLDSVILNGYAFSDRYDLAMSKNNKYSQDLANAIKGLQSAKIKVLSLDVLPNOLYNLP 925

Db 877 QFLDSVIQNGYAFTRDYDLGMSKANKYGTADQLVKAIKALHAKGLKVMADWVPDQMYTFP 936
QY 926 GKVVTAFRVNOYQAKSGATINTKPYVANTRSYG-DYQEOYGGKFLDDLOKLYPRLPST 984
DB 937 KQEVVTVTRTDKFGKPIAGSQINHSLYVTDTKSSGDDYQAKYGGAFDLDELKEXPELFTK 996
QY 985 KOISTGKPIDPSVKITNWSAKYFNGSNLGRGAKYVLSE--GNKYNLNLADGKLPLPTVLN 1042
DB 997 KOISTGKPIDPSVKITNWSAKYFNGSNLGRGAKYVLSE--GNKYNLNLADGKLPLPTVLN 1056
QY 1043 NTYGOPOVSANGFISKNGGHIYLDKNGOEVRKFRKEISGSWYFYFSDGKMATGKTIGND 1102
DB 1057 GKVESGIRYDG---KGYIYNSSATGDQVAKASFITEAGNLNLYFGKDGMYVTGATINGA 1112
QY 1103 TYLFMPNGKQLKEGVWVDGK-KAYYYDDNGRTWTKGVPFRVDGQDKRWYFNGDGTIAI 1161
DB 1113 NYFFLENGTALRNITTYDAQNSHYANDGKRYEN-GYQF---GND-WRYFK-DGNMAV 1166
QY 1162 GLVSLD----- 1167
DB 1167 GLTTVDGNNVOYFDKDGVOAKDKIIIVTRDGKVRVYFDQHNNGNAATNFTIADKTHWYVLGD 1226
QY 1168 -----NRTLYFDAYGYQVKGOTVTIN-GKSYTFDADQDGL----- 1201
DB 1227 GVAVTGAQTVGKQKLYFEANGQVKGDFVTSDEGKLYFYDVSQDMWTDTFIEDKAGNWF 1286
QY 1202 ----- 1201
DB 1287 YLKGDAAVTGAQTIRGQKLYFKANGQVKGDIIVKGTGKIRYYDAKSGEQVFNKTVKAA 1346
QY 1202 -----VOTDNANPAP----- 1231
DB 1347 DGKTYVINGDGVADVPSVVKQOTFKDASGALRFLYNLKGQVLTGSGWYETANHDWYIOSG 1406
QY 1232 QLLTGEQITDQGVFFQDNGVQVKGATATDASGLVRFYDRDQGHQ----- 1276
DB 1407 KALIGEQTINGQHLVYFREDGHQVKGQVLTGDKVRVYDANSQDQAFNKSVTVNGKTYYP 1466
QY 1277 -----VKGWYSTSDDDNWNVYVNSGQVLTGLQ 1303
DB 1467 GNDGTAQTAGNPKGQTFKDGSDIRFYSMEGQVLTGSGWYENAGQWLYV-KNGKVLTLGLQ 1525
QY 1304 TIDQTYVYFDDKQKQKAVWDENGLRYFDADSGNMLRDRKWNVDGWNWYFNRNG 1360
DB 1526 TVGSRVYFDENGIQAKGKAVRTSDGKIRYFDENSGSMITNQWKFVYGOYIYFGNDG 1582

RESULT 14
ADX37280
ID ADX37280 standard; protein; 1590 AA.
XX
AC ADX37280;
XX 21-APR-2005 (first entry)
XX
DE Streptococcus sobrinus glucan binding protein B.
KW immunogenicity; immune stimulation; glucan binding protein-B;
KW microparticle; major histocompatibility complex; tooth disease.
XX Streptococcus sobrinus.
XX US2005031633-A1.
XX 10-FEB-2005.
XX 09-MAR-2004; 2004US-00797821.
XX 13-APR-1998; 98US-0081550P.
PR 08-JAN-1999; 99US-0115142P.
PR 12-APR-1999; 99US-0029004P.
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 19:39:30 ; Search time 29.8857 Seconds
(without alignments)
3776.130 Million cell updates/sec

Title: US-10-797-821-39
Perfect score: 7230
Sequence: 1 MEKNRLRYKLHKYKQWVAIG.....MKNVNDWNYFNRNGLATRW 1365

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA*
1: /cgn2_6/ptodata/1/1aa/5 COMB.pdp.*
2: /cgn2_6/ptodata/1/1aa/6 COMB.pdp.*
3: /cgn2_6/ptodata/1/1aa/H COMB.pdp.*
4: /cgn2_6/ptodata/1/1aa/pCTUS COMB.pdp.*
5: /cgn2_6/ptodata/1/1aa/RE COMB.pdp.*
6: /cgn2_6/ptodata/1/1aa/backfiles.pdp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	3174	43.9	1430	2	US-09-008-172-2
2	3174	43.9	1430	2	US-09-210-361-6
3	3174	43.9	1430	2	US-09-740-274-6
4	3026.5	41.9	1577	1	US-08-793-824-2
5	2945.5	40.7	1475	2	US-09-007-999-2
6	2945.5	40.7	1475	2	US-09-210-361-2
7	2945.5	40.7	1475	2	US-09-740-274-2
8	2893.5	40.0	1375	2	US-09-210-361-4
9	2893.5	40.0	1375	2	US-09-740-274-4
10	2227	30.8	2057	2	US-09-499-203-2
11	2191.5	30.3	1278	2	US-09-604-957-3
12	2191.5	30.3	1781	2	US-09-995-749A-2
13	1635.5	22.3	545	2	US-09-604-957-4
14	1597.5	22.1	545	2	US-09-995-749A-10
15	1362.5	18.8	523	2	US-09-604-957-5
16	1361	18.8	522	2	US-09-995-749A-11
17	1203.5	16.6	535	2	US-09-604-957-7
18	1203.5	16.6	535	2	US-09-995-749A-13
19	1173	16.2	584	2	US-09-604-957-6
20	1171	16.2	584	2	US-09-995-749A-12
21	450	6.2	349	2	US-09-009-620-2
22	365.5	5.1	2710	1	US-08-480-604A-6
23	365.5	5.1	2710	1	US-08-405-496A-6
24	365.5	5.1	2710	2	US-08-915-136-6
25	365.5	5.1	2710	2	US-08-957-310-6
26	365.5	5.1	2710	2	US-10-011-366-6
27	365.5	5.1	2710	2	US-09-084-517-6

28	330.5	4.6	1231	2	US-08-714-741-41	Sequence 41, Appl
29	327.5	4.5	2366	1	US-08-480-604A-10	Sequence 10, Appl
30	327.5	4.5	2366	1	US-08-405-496A-10	Sequence 10, Appl
31	327.5	4.5	2366	2	US-08-915-136-10	Sequence 10, Appl
32	327.5	4.5	2366	2	US-08-957-310-10	Sequence 10, Appl
33	327.5	4.5	2366	2	US-10-011-366-10	Sequence 10, Appl
34	327.5	4.5	2366	2	US-09-084-517-10	Sequence 10, Appl
35	323	4.5	866	2	US-09-545-773-2	Sequence 2, Appl
36	323	4.5	866	2	US-10-222-038-2	Sequence 2, Appl
37	318.5	4.4	811	1	US-08-480-604A-7	Sequence 7, Appl
38	318.5	4.4	811	1	US-08-405-496A-7	Sequence 7, Appl
39	318.5	4.4	811	2	US-08-915-136-7	Sequence 7, Appl
40	318.5	4.4	811	2	US-08-957-310-7	Sequence 7, Appl
41	318.5	4.4	811	2	US-10-011-366-7	Sequence 7, Appl
42	318.5	4.4	811	2	US-09-084-517-7	Sequence 7, Appl
43	318.5	4.4	812	1	US-08-480-604A-29	Sequence 29, Appl
44	318.5	4.4	812	2	US-08-915-136-29	Sequence 29, Appl
45	318.5	4.4	812	2	US-09-084-517-29	Sequence 29, Appl

ALIGNMENTS

RESULT 1
US-09-008-172-2
; Sequence 2, Application US/09008172
; Patent No. 6127602
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0358D
; CURRENT APPLICATION NUMBER: US/09/008,172
; CURRENT FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-008-172-2

Query Match	43.9%	Score 3174;	DB 2;	Length 1430;
Best Local Similarity	47.3%	Pred. No. 5.2e-217;		
Matches	677;	Conservative 213;	Mismatches 436;	Indels 104; Gaps 38;
QY	1	MEKNRLRYKLHKYKQWVAIGVTT--VTLSFLAGGVAAADTNNNDGTSV-----QVNMKV	53	
Db	1	METKRRYKMHKXKHVTVAVASGLITLTGSSVSAETEQOTSQSDKVVTKSEDDKAA	60	
QY	54	PSDPKFDA-----QAQNGQLAAMFKAANOADOTATQSVSPATDGRVDNQVTPAANQPAAN	109	
Db	61	SESSQTDAPKTQAOQTEQ-TQAQ-SQANVAD-TSTS-----ITKETPSQNIITTAQNSDDKT	113	
QY	110	VANQDVANPATDAGALNQSAADTSDCKAVP-----QTSDFG-----H	149	
Db	114	VTNTKSEAOISETRTKQSEAOITASSQALTKQAKAELTKQRTAQENKPNVDLAAIPN	173	
QY	150	LETVDGKTYTVDANGQRLKNTSMVIDGTYTFDGTGEAQTDLP-----KTQANDQNDVPD	205	
Db	174	VKQIDGKYYIGSDGQPKNFALTNNKVLFDKNTG-ALTDTSQYQFQGLTKLNN----	229	
QY	206	SYQANNQAYSNEASFFETVDNYLTADSWYRPRKILKNGSQWASSEGDLRPLMTWPDPA	265	
Db	230	DYTPHNQIVNFEITSLETIDNYVTADSWYRPRKILKNGKTWTASSESDLRPLLMWSWPK	289	
QY	266	ATKAAAYNFWAKEGLISG-SYRQNSA--NLDAANTQIOAIEKKIASSEGTNTWLRDKMSQ	322	
Db	290	QTQIAYLNMNQGLGTGENTADSSQESLNLAATQVQVKIETKISQTOQTQWLRLDINS	349	
QY	323	FVKSQNQSWSIASENETVYPNQDHMQGALLFSNSKDTSEHANSWMLLRNRPFTQTGKQKY	382	


```
QY 553 RIAKIIRKLGKTNAGLNLTLDDLNKAFDIYNQDMNATDKVYPPNNLPMAAYMWLQNK 612
Db 590 VIAKIIKAQI-NPKTDGLT-FTLDELKQAFKIYNEQDMRQAKKKYQTSNIPTAYALMSNK 647
QY 613 DVTETVYVYGDMDYDNGQYMATKTPFNALETLLKGRIKYVAGGO--AVSYKQ-----DW 664
Db 648 DSITRLYYGDMYSDGQYMATKSPYDAIDTLKARIKAYAGGQDMKITYVGGDKSHMDW 707
QY 665 S-SGLITSVRYGKANSADAGNTETRNCGMALLNNRPNFRAYRN--LTLNNGAAHKSQ 721
Db 708 DYTGLVTSVRYGTGANEATDQSEATKTQGMVAVITSNNPSLKNQNDKVIYNNGAAHKNQ 767
QY 722 AYRPLLLSTKGIATYLNDSVDVSQYKYTDQGNLSFASLSQSVANAQVSGMIQVWYP 781
Db 768 EYRPLLLTKOGLTSYTSDAAKS-LYRKTNDKGLVFDASDIQGYLPQVSGYLAVWVP 826
QY 782 VGAAADNQVRTSPSQAATKGNHLYHQSDALDSOVIYEGFSFOAQAQSPDQYTNVIAKN 841
Db 827 VGASDNQDVRVAASNAKANATQVYESSALDSQLIYEGFSNFDQVFTKDSYTNKKIAQN 886
QY 842 GDLFKSGWITQPEMAPQVYSSBDGTFLDSVILNGYAFSDRYDLAMSNNKYSGKODLANA 901
Db 887 VQLFKSGWTSFEMAPQVYSSBDGSLDSIIQNGYAFEDRYDLAMSNNKYSGQODMINA 946
QY 902 IKGLSAGIKVLSLVPNLQNLNPKGVVTVATRVNQYQAKSGATINKTPYVANTRSYG- 960
Db 947 VKALHSGIQVIAWVPDQIYNLPKGVVTVATRVNDYGEYRKDSEIKNTLYAANTKSNKG 1006
QY 961 DYQEOYGGKFLDDLQKLPRLPSTQIISTGKIDPSVKITNWSAKYFNGSNILGRGAKYV 1020
Db 1007 DYQAKYGGAFJSELAAPYSPISFNRTQISNGKKIDPSEKITAWEKAKYFNGTNIILGRGVY 1066
QY 1021 LSE--GNKYLNLADGKLPVTLNNTYQPOVQVANGFTISNGGIIHYLDKNGQEVKNRF-K 1077
Db 1067 LKDNASDKYFELKGNQVILPKQNTN-----KEASTGFVNDGNGMTFYSTSGYQAKNSFVQ 1121
QY 1078 EISGSWYFYDSDGKMATKTKIGNDTYLFMPNGKQKKEGVW--YDGKAYYYDDNGRTWT 1135
Db 1122 DAKGNWYFYDNGHVMYGLQQLNGEVQYFLSNGVQLRSLFLENADGSK-NYFGLHNGYS 1180
QY 1136 NKGFEVFRVGDQKRYFNGDGTIAI GLVSLDNRTLYFDAYQYKVGQTVT-INGKSYTF 1194
Db 1181 N-GYTSF--DNDSKRYFVDSAGVMAVGLKTINGTQYFDQDQYQVKGAWITGSDGKRYF 1237
QY 1195 DADQDLVQTDNANPAPQOAGKLLGNQWGY-RKDGOLLTGEOTIDGQKVFQDNGVQ 1253
Db 1238 DSGSGMAVNPANDK-----NGDMWYLSNDGIALVGVQTINGKTYIFQDQKQ 1286
QY 1254 VKGQTATDASGLYFRDQGHQVKGWYSTSDNNWYVNBESGQVLTGLQITIDGQTVYFD 1313
Db 1287 IKGKIITD-NGKLYFLANSBELARNIFATDSQNNWYFGSDGVAVTSQTIAGKLIYA 1345
QY 1314 DKGQAKKAVWDENGNRYFDADSGNMLRDKW-NVDGNWYFNRNGLA 1362
Db 1346 SDGQVKGFSV--TYNGKVHYHADSGELQVNRFEADKQGNWYLYDSNGEA 1394
```

RESULT 3

```
US-09-740-274-6
; Sequence 6, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
```

```
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-6
```

```
Query Match 43.9%; Score 3174; DB 2; Length 1430;
Best Local Similarity 47.3%; Pred. No. 5.2e-217;
Matches 677; Conservative 213; Mismatches 436; Indels 104; Gaps 38;
```

```
QY 1 MEKNRYKLHKYKQWVAIGVTT--VTLSFLAGGVVAADTNNNDGTSV-----QVKNV 53
Db 1 METKRYKHKYKQWVAIGVTT--VTLSFLAGGVVAADTNNNDGTSV-----QVKNV 53
QY 54 PSDPKFDA----QAQNGQLAQAMFKAANQADQATATSOVSPATDGRVDNQVTPAANQPAAN 109
Db 61 SESSQTDAPKTQAQTEQ-TQAQ-SQANVAD-TSIS----ITKETPSQNIITQANSDDKT 113
QY 110 VANQVANPATDAGALNRQSAADTSTDGKAVP-----QTSDDPG-----H 149
Db 114 VNTKSEEAQSEERTKQSEEAQTTASSQALTKQAKAELTKQRTAAQENKPNVDLAAIPN 173
QY 150 LETVDGKTYVDANGQRKKNYSWIDGKTYYPDGTGGAQTDLP----KTQGANQDNVDP 205
Db 174 VKQIDGKTYIIGSDGQPKKNFALTNNKVLVYDKNTG-ALTDTSQYQFQGLTKLNN--- 229
QY 206 SYQANNOAYSNEASFEVDNLTADSWYRPRKILKNGSQWASSEGDLRPLMTWVPA 265
Db 230 DYTPIHQIVNENTSLTIDVNTADSWYRPRKILKNGKTWTASSEDRLRPLMSWVDPK 289
QY 266 ATKAAYANFWAKEGLISG-SYRQNSA--NLDAATONIQAIEKKIASEGNTNWLDRKMSQ 322
Db 290 QTQIAYLNMNQGLGTGENTYADSSQESLNLAQTVQVKIETKISQTOQTQWLRIINS 349
QY 323 FVKSQNQSIASENETVYPNODHMOGGGALLFESNKTDEHANSDWELLNRPFTQKQKY 382
Db 350 FVKTPQNNWSQTESDTSAGEKDLQGGALLYSNDSKTAYANSDYRLNRTPTSQTGPKY 409
QY 383 FTTNYA-GYELLANDVDSNPVQAEQLNHLHYLNMWGDIVMGDKDANFDGVRVDAVN 441
Db 410 FEDNSSGGYDFELLANDIINSPVQAEQLNHLHYLNMYSIVANDPEANFDGVRVDAVN 469
QY 442 VNADLLQRTDYKAKYCTQNEKNVADHLSILEAWSGNDNDYVQDQNNFSLINDQRS 501
Db 470 VNADLLQIASDYLKAHYGVKSEKNAINHLSILEAWSNDPQYNKDTKGAQLPIDNKLRL 529
QY 502 GMLKAF-----GYASAYRGNLSNLATAGLKNRSA-NPDSDPVNPVYFIRAHSEVOT 552
Db 530 SLLYALTRPLEKQDASNKNEIRSGLEPVITNSLNNRSAGEKNSERMANVYFIRAHSEVOT 589
QY 553 RIAKIIRKLGKTNAGLNLTLDDLNKAFDIYNQDMNATDKVYPPNNLPMAAYMWLQNK 612
Db 590 VIAKIIKAQI-NPKTDGLT-FTLDELKQAFKIYNEQDMRQAKKKYQTSNIPTAYALMSNK 647
QY 613 DVTETVYVYGDMDYDNGQYMATKTPFNALETLLKGRIKYVAGGO--AVSYKQ-----DW 664
Db 648 DSITRLYYGDMYSDGQYMATKSPYDAIDTLKARIKAYAGGQDMKITYVGGDKSHMDW 707
QY 665 S-SGLITSVRYGKANSADAGNTETRNCGMALLNNRPNFRAYRN--LTLNNGAAHKSQ 721
Db 708 DYTGLVTSVRYGTGANEATDQSEATKTQGMVAVITSNNPSLKNQNDKVIYNNGAAHKNQ 767
```

Qy	722	AYRPLLLSTKDGATATYLNDSGDSRQYKYTDSQGNLSFSASELSQSVANAQVSGMIQWVP	761
Db	768	EYRPLLLTTKDGTSYTSDDAAKS--LYRKNTDKGELVFDASDIQGLNPNQVSYLAWVP	826
Qy	782	VGAADNQDVRTSPSTQATKGNITYHOSDALDSQVIVEGFSNFOAFAQSPODYTNVIAKN	841
Db	827	VGASDNQDVRVAASNKANATQVYESSALDSQLIYEGFSNFQDFVTKUSDYTNKKIAQN	886
Qy	842	GDLFKSGWITQFEMAPQVYSSSDGTFLDSVILNGYAFSDRYDLAMSKNNKYGSKQDLANA	901
Db	887	VQLFKSGWVTSFEMAPQVYSSSDGSLDSIIQNGYAFEDRYDLAMSKNNKYGSOQDWINA	946
Qy	902	IKGLQSGIKVLSDLPVNOIYNLPKGKVVWATATRVNQYQOAKSGSATINKTTPYVANTRSYG-	960
Db	947	VKALHKSIGIQVIAADWPDQIYNLPKGKVVWATATRVNDYGEYRKDSEIKNTLYAANTKSNKG	1006
Qy	961	DYQEOYGGKFLDDLQKLYPLRFLSTKQISTGKPIDPSVKIITNWSAKYFNGSNILGRGAKYV	1020
Db	1007	DYQAKYGGAFLSLAAKYPSIFNRTQISNGKKIDPSEKITAWKAKYFNGLILGRGVGYV	1066
Qy	1021	LSB--GNKYLNLDAGKULFPLTVLNTVYGPQVQVANGFISKNGGIHYLDKNGQEVKQRF-K	1077
Db	1067	LKDNASDKYPELKGNOTYLPKQMTN-----KEASTGFVNDGNGMTFYSTSGYQAKNSFVQ	1121
Qy	1078	EISGSWYYPDSDGKMATKTKIGNDTYLVFWPNGKQOLKEGYW--YDGKCAYYVDNNGRTWT	1135
Db	1122	DAGNWTYFPNNGHMVYGLQOLNGEVQYFUSNGVQLRESFLENADSK-NYFCHLGNRYS	1180
Qy	1136	NKGFEVPRVDGQKWRYFNGDGGTIAIGLVSLDRNTLYFDAYGQVKGQTVT-INGKSYPF	1194
Db	1181	N-GYSPF--DNDSKWRYFDASGVMAVGLKTINGNTQYFDQDGYQVKGAWITGSDGKKRYF	1237
Qy	1195	DAGQDLVQTDNANPAPQOAGKWLGDNOWGY-RKDGQLLTGEOITDGQKVFQDNGVQ	1253
Db	1238	DDGSGNNAVRFANDK-----NGSDWYLYNSDGIALVGVQTINGKTYTFYFGDQKG	1286
Qy	1254	VKGGTATDASGVLFYDRDGHQVKGWYSTSDDNWYVNESGOVLTLGLQITDGOVTFD	1313
Db	1287	IKGKIITD-NGKLYFLANSGBELARNIPATDSQNNWYFSGDSGVAVTGSQTIAGKKLYPA	1345
Qy	1314	DKGIQAKGVAWBENGNLRYFDADSGNMLDRWK-NVDGNWYYPNRRGLA	1362
Db	1346	SDGKQVKGSGFV-TYNGKVHYTHADSGLOVNRFEADKGNWYLYDSNGEA	1394

RESULT 4

[illegible]

```
QY 654 GQAVSYKQWSGILTSVRYGKANSASDAGNTETRSNGMALLINNRPNFRAYRN--IT 711
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 712 LNMGAHKSQAVRPILLSTKDGIAATVLDSDVDSQYKYTDQGNLSPSASELSQVANAQ 771
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 834 VNMGAHKSQAVRPILLSTKDGIAATVLDSDVDSQYKYTDQGNLSPSASELSQVANAQ 893
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 772 VSGMIOVWVYVGAADNQDVRTSPSTQATKDG--NIYHQSADLDSQVIYEGFSNFOAASP 830
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 894 VSGYLAVWVYVGAADNQDVRTSPSTQATKDG--STKKGQVFPSSAALDSQVIYEGFSNFOAASP 951
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 831 DOYTNVATKNGDLKPSGICITOFEMAPQVYSSDGTFLDSVLNGYAFSDRYDLAMSKN 890
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 952 SOYTRVIAQNAKLFKMGWITFSEFAPQVYSSDGTFLDSVLNGYAFSDRYDLAMSKN 1011
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 891 KYGSKODLANATKGLQSGIKVLSLVPNQLNLPKGVVATRVNOYQAKSGATINKT 950
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1012 KYGSKODLANATKGLQSGIKVLSLVPNQLNLPKGVVATRVNOYQAKSGATINKT 1071
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 951 PYVANTRSYG--DYQBYGGKFLDDQLKLYPRILFSTKQISTGKPIDPSVKITNWSAKYFNG 1009
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1072 LYAAKTRTFNGDFQKYGGAFDELKAKYPAIFERVQISNGRKLTTNEKITQWSAKYFNG 1131
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1010 SNILGRGAKYVLSG--NKYLMADGKFLPLTVLANTYQPOQVANGFISKNGGHIYLDK 1067
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1132 SNIQGTGARYVLQDNATNQYFSKAGQTFPKQMT-----EITGSGFRRVGDVQYLSI 1185
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1068 NQEVKNRFEKI--SGSWYFSDGKMATCKTKIGNDTYLFMPNGHQLKEGV----- 1117
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1186 GGYLAANTFIOVGANQWYFEDKGNMVTGEQVIDGKKYFFLDNGQLRHVLRQSGDHVY 1245
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1118 WYD--GKXAY-----YDDNGRTW-----TNKGFVE 1141
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1246 YVDKGVQAFNGFYDFAGPRODVRFDGNGQMYRGHDMYGTTFYFDEKGTGQAKDKFIR 1305
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1142 FRVGDQK-----WRYFNGDGTIAIGLVSQNLRTLYFDAYGYQ 1179
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1306 F-ADGRTYFIPDTGNLAVNRFAQNPENKAWYLDNSGYAVTGLQTINGKQYVFDNEGRQ 1364
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1180 VKGQVTLTKNGSYTFDADQGLVQ---TDNAPAPQOAGKLLGDNQWGY--RKGQGLL 1234
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1365 VKGHFTVNNORYFLDGDGSEIAPSRFVTEN-----NKWYVVDGNGKLV 1408
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1235 TGEQITDQKQVFFQDQVGGTATDASGLVRFYDRDQHQVKGWYSTSDNNVYNE 1294
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1409 KGAQVINGNHYYFNNDYQVKGAWANG-----RYDGSQAVSNQFTQIAANQWAYLNQ 1463
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1295 SGQVLTGLQITDQVTFYFDDKGIQAKGKAVMDENGLNRYFPDADSGNMLDRW--KNVDGNW 1353
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1464 DGHKVTGLQINNKVYVFGSNGAQVKGKLL--TVQGGKCYFDAHTGEQVVRPVEAARGCW 1522
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1354 YFNRGLA 1362
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1523 YFNSAGQA 1531
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 5

```
US-09-007-999-2
; Sequence 2, Application US/09007999
; Patent No. 6087559
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
;   TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0356D
; CURRENT FILING DATE: 1998-01-16
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
```

; LENGTH: 1475

; TYPE: PRT

; ORGANISM: Streptococcus mutans

US-09-007-999-2

Query Match 40.7%; Score 2945.5; DB 2; Length 1475;

Best Local Similarity 45.8%; Pred. No. 1.1e-200;

Matches 652; Conservative 225; Mismatches 425; Indels 121; Gaps 40;

QY 1 MEKILRYKHLKVKQKQWVAIGVTT--VTLSFLAGGQVVAADTN-----NNDGTSVQVVK 51

Db :

QY 1 MDKVKRYKLRKVKKRWVTVSVASAVMTLTLSGG--LVKADSNEKSKQISNDSTSV-- 55

Db :

QY 52 MYPDPKFDQAQNGQLAQAMFK---AANQADOT-----ATSOVSPATDGRVDNQTTPA 102

Db :

QY 56 -----VTANESNVITEATSKQEAASSQTNHTVTTSSTSVVNPKE--VVSNPYT-- 104

Db :

QY 103 ANQPAAN---VANQDVANPATDAGALNRQSAADTSTDGKAVP-----QTSQDPQHLET 152

Db :

QY 105 VGETASNGEKLQNTTVDTKTSEAAANNISKOTTEADTDVIDDSNAANLQILEKLPNVE 164

Db :

QY 153 VDGKTYVVDANGORLKNYSWIDGKTYVFDGOTGEAOTDLPKTGOANQDNV---PDSYQA 209

Db :

QY 165 IDGKTYVVDNNGKVRNTFTLIADGKILHFD--ETG--AYTD--TSIDTVNKDITVTRSNLYK 221

Db :

QY 210 NNQAYSNEASPFETVDNYLTADSWYRPRKILKNGSQWQASSEGDLRPIMLTWMPDAATKA 269

Db :

QY 222 YNQVYDRSAQSFHVHYLTAEWYRPKYLLKDGKTWTQSTEKDFPPLMTWMPDETQ 281

Db :

QY 270 AYANFWAKEGLISGY--RONSANLDAATQNTQSAIEKKIASSEGNNTNWLDRKMSQFVK 327

Db :

QY 282 QVYVYMAQLGINKTYDDTSNQLNIAAATQAKIEAKITTLKNTDMLRQTISAFVK 341

Db :

QY 328 NQWSTASENEITYPNQDHMOGALLFSN--SKDTEHANGSWRLNRPPTQCGQ--KYFT 384

Db :

QY 342 SAWNDSSEK---PPDHLQNGAVLYDNEGKLTPTYANSNYRLNRPPTQCGKDPRTYA 397

Db :

QY 385 TN--YAGYELLANDVDNSNPVQABQLNHLHYLMNWGDIVMGDKDANFGVVDVADNVN 443

Db :

QY 398 DNTIGGYEFLANDVDNSNPVQABQLNHLHYLMNWGDIVMGDKDANFGVVDVADNVN 457

Db :

QY 444 ADLQIQDQYKAKYGTQDNKNAIDHLSILEAMSGNDYVYKQDNFSLSDINQORS 503

Db :

QY 458 ADLLQIAGDYLKAAKGIHKNDKAAANDHLSILEAMSGNDYVYKQDNFSLSDINQORS 517

Db :

QY 504 LKAFYASAYRGNLSNLTATAGLKRS--ANPDSDPVNPVYFTRAHDSEVQTRAKIIREKL 562

Db :

QY 518 LFSLAKPLNQSRGWNPLITNSLVNRTDDNAETAAPVPSYFTRAHDSEVQDLDIADIIKABI 577

Db :

QY 563 GKTNADGLTNLTDDLNKAFDIYNQDMNATDKVYVPPNLPWAYAWMLQNKDTVTVRYGD 622

Db :

QY 578 -NPVVGY--FTMEIKAFIYNKDLATEKKYTHYNTALSYALLLNKSSVPRVYGD 635

Db :

QY 623 MYDNGQYMATKTPFYNAIETLLKRIKYVAGGQAVSYKQDWSSGILTSVRYGKANSAS 682

Db :

QY 636 MFTDDGQYMAHKTINYBAIETLLKARIKYVSGQAMRNQVGNSEIITSVRYGKALKAT 695

Db :

QY 683 DAGNTEFRNSGMALLINNRPNFR--AYRNLTNMGAAHKSQAYRPLLSTKDGIAATYND 740

Db :

QY 696 DTGDRTRTSGVAVIEGNNPSLRKASDRVVVNMGAHKSQAYRPLLSTKDGIAATYND 755

Db :

QY 741 SDVDSQVYKTYDSCNLSFSAELQSVANAQVSGMIOVWVPGAA-----DNQD 789

Db :

QY 756 QEA--AGLVRYTNDRGELIFTAADIKGYNAPQVSGYLGWVVPVGAALIKNFALRLARPHQ 814

Db :

QY 790 VRTSPSTQATKDGNIYHQSADLDSQVIYEGFSNFOAASPQDQYTNVIAKNGDILFKSWG 849

Db :

QY 815 MAS-----VHQNAALDSRVMEFEGFSNFOAFATKKEEYTNVIAKNDVDFEAWG 862

Db :

QY 850 ITQFEMAPQVYSSDGTFLDSVLNGYAFSDRYDLAMSKNKNYKSGQDLANAIKGLQASG 909

Db :

QY 863 VTFEMAPQVYSSDGTFLDSVLNGYAFSDRYDLAMSKNKNYKSGQDLANAIKGLQASG 922

Db :


```

; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
; US-09-740-274-4

Query Match      40.0%; Score 2893.5; DB 2; Length 1375;
Best Local Similarity 44.3%; Pred. No. 4.9e-197;
Matches 634; Conservative 211; Mismatches 412; Indels 175; Gaps 30;

QY 1 MEKRLRYKLHKVKKQWALGVTT--VTLSELAGGVVAADTNNNGTSTVQVKNMVPSPDK 58
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MEKRVRFKLKVKKRWVTVSIAVVTLTSL--GSLVKADSTDDRQAVT-----49

QY 59 PDAQQNGQLAQMFKAAQAD--QTATSOVSPATDGRVDN--QVTPAANQPAANQDVA 116
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
50 -ESQASLVTTSEAKETLTATDTSTATSTATSTOPTATVDNVSTNQSTNTANTAN--FVV 107

QY 117 NPA-----TDAGALNRQSAADTSTDCGKAVPQTS-----145
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
108 KPTTTSEQAKTNSDKIITTSKAVNLTAT----GKFPVANNNTAHPKTVTDKIVPIKP 162

QY 146 -----QPG-----HLETVDGKTYVDVANDQORLNKNSWIDGKTYFPDQGT 185
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
163 KTGKLPQPSLSQDDIAALGNKRNKNGKTYYYKEDGTQKNYALNNGKTFPFDGTG 222

QY 186 GEAQTDLR--KTQGANQDNVPSVQANNOAYSNEASSFETVDNYLTADSWYRPRKILKNGQ 244
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
223 ALSNNTLPKKGNIITNDNTNSFAQYNQYSTVDVANFEHVDHYLTAEWSYRPRKYLKDGK 282

QY 245 SQAASSEGDLRPLMTWPDAAATKAAYANFWAKEGLISGSYRQNSA--NLDAAATQNIQSA 302
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
283 TWTQSTEXDIFRLLMTWPDQETQRYVYNNYNAQLGIHOTYNTATSPILQNLAAQTIQTK 342

QY 303 IEKKTASEGNTWLRDKMSQFVKSQNSQWSIASENETVYPNQDHMQGALLFS--NSKOTEH 361
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
343 IBEKITAENKTNLRTQTSIAFVKTQSAMNSDSEK----PFDDHLQKGLLYSNNKSLTSQ 398

QY 362 ANSDRLNRNPTFTGKQ--KYFT--TNYAGVELLNDVNSNPVQAEQLNHLHYLMN 418
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
399 ANSNRYILNRTPTNGKDKPRYTADRTTGGYEFELLANDVDNSNPVQAEQLNHLHYLMN 458

QY 419 WGDIVMGOKDANFDGVRDADVNVDNLQIQRDYVYKAKYGTQDQNEKNAIDHLSLEAWS 478
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
459 FGNIVANDPDANFDSIRVDADVNDVADLLQIAGDYLKAAKGHKKNDKAANDHLSLEAWS 518

QY 479 GNDNDYVKDQNNFSLINDQSRGMLKAFGYASAYRGNLSNLATAGLKNRS--ANPDSDPV 537
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
519 YNDTFYLHDDGDMINMDNRLSLLSYSLAKPLNQSGMNPLITNSLVNRTDODNAETRAV 578

QY 538 PNYVIRAHDSVQTRIAKIREKLGKTNWADGLTNLTDLLNKAFDINQDMNATDKVY 597
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
579 PSYSPIRAHDSVQDLIRNIIRTE--NPNVVGY--SFTTEIKKAFEYNNKOLLATEKKYT 636

QY 598 PNNLPMAYAMWLQNKDTRVYVYDGYMDNGQYMATKTPFFYNAIETLLKGRIKYVAGQA 657

```

RESULT 10

```

US-09-499-203-2
; Sequence 2, Application US/09499203
; Patent No. 6570055
; GENERAL INFORMATION:
; APPLICANT: KOSSMANN, Jens
; APPLICANT: WELSH, Thomas
; APPLICANT: QUANZ, Martin
; APPLICANT: KNUTH, Karola
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternansucrase
; FILE REFERENCE: 147-196P
; CURRENT APPLICATION NUMBER: US/09/499,203
; CURRENT FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2057
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-09-499-203-2

```

```

Db 637 HYNLTALSYALLLTNRKSSVPRVYVYGMFTDDGQYMAHKTINYBAIETLLKARIKYVSGQA 696
QY 658 VSYKQDWSSGILTSVRYGKGANSASDAGNTEFNSGMALLINRPNFR--AYRNLTLNMG 715
Db 697 MRNQQVGNSEIITSVRYGKGALKATDGTDRTRTSGVAVIEGNNPSRLKASDRVVNMG 756
QY 716 AAKHSQAYRPLLLSTKDGATYLNDSVDSDRQVKYTDSCNLSFSASELSQSVANAQVSGM 775
Db 757 AAKHQAYRPLLLTTDNGIKAVHSQEA--AGLVRYTNDRGELIFTAADIKGANPOVSGY 815
QY 776 IOVWVPVGAADNQDVRTSPSTQATKGNIIYHQSDDLSQVIYEGFSNFOAFQSPQYTN 835
Db 816 LGVWVPVGAADQDVRVAASTAPSTDGKSVHQNAALDSRVMPFEGFSNFOAFATKKEEYTN 875
QY 836 AVIAKNGDLFKSGITQFEMAPQVYSSDGTFLDSVILNGYAFSDRYDLAMSKNKYKSK 895
Db 876 VVIKRVKFAEWGVTDFEMAPQVYSSDGTSGLSVIQNGYAFTRDYDLGISKPNKYGTA 935
QY 896 QDLANAIGLOSAGIKVLSDLVPNOLXNLPKGEVVTATRVNOYQOAKSGATINKTPVAN 955
Db 936 DDLVKAIKALHSGIKVWADWVPDQMYALPEKEVVTATRVDKYGTVPAGSOIKNTLYVVD 995
QY 956 TRSYG--DYQEQYCGKFLDDLQKLYPRLPSTKQISTGKPIDPSPVKITNWSAKYPNGSNILG 1014
Db 996 GKSSGKQOQAKYGGAFLEELQAKYPPELPARKQISTGVPMDPSVKIKQWSAKYFNGTNILG 1055
QY 1015 RGAKVYVLS--GNKYLNLADGKLFPLTVLNNYTGPOVSANGFISKNQGIHILDKNGQEV 1072
Db 1056 RGAGVYLKQDQATNTYFSLVSDNTFLPKSLVNPNGHTSSSVTGLVFDGKGVYVYSTSGNQA 1115
QY 1073 KNRFEISGSWYTFSDGKMATGKTIGNDTYLFNPNGKQKGEVYVYDCKKAY--YYDDNG 1131
Db 1116 KNAFISLGNWYTFDNGYMWVTAQSGINGANTYFISNGIQLRNALYDNGKVLSTYNGNDG 1175
QY 1132 RTWTKGFVFRVDCQKWRFYENGDTTIAIGVLSLDNRTLYFDAYGYQVQKGTVTINGKS 1191
Db 1176 RRYEN-GYVLF--CQ-QWRYFQ--NGIMAVGLTRVHGAQVYFDASGFOAKGQFI-----1223
QY 1192 YTFDADQDGLVQTDNANPAQOQAGWKLKLDNQWGYRKDQGLLTGEQITDGKVPFDNG 1251
Db 1224 -----1223
QY 1252 VQVKGTTATDASGLVRFYDRDQGHQVKGWYSTSDNNWVYVNESGOVLTLQOTIDGQTVY 1311
Db 1224 -----TTADGKLYFDRDSGNQISNRFRVNSKGWFLFDHNGVAVTGTVTFFNGQLY 1275
QY 1312 FDDKGIQAKGAVMDENGLRYFDADSGNMLRDW--KNVDGNWYFYFNRLGLA 1362
Db 1276 FKPNGVQAKGEFIRDANGLYRYDPSNGNEVRNFRVNSKGWFLFDHNGIA 1327

```



```
; SEQ ID NO 3
; LENGTH: 1278
; TYPE: PRT
; ORGANISM: Lactobacillus reuteri
US-09-604-957-3

Query Match      30.3%; Score 2191.5; DB 2; Length 1278;
Best Local Similarity 40.9%; Pred. No. 48e-147;
Matches 518; Conservative 170; Mismatches 436; Indels 141; Gaps 34;

QY 36 AADTNNNDGTSV--QVKN-MVPSDPKFDQAQNG--OLAQAMFKAANOADQTATSOVSPAT 91
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 78 AADGNSGDNVTVDQFSKNYATTGNGFDYVKVNGNQVFSGWHATNQSNDKDSQWIIIVLV 137
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 92 DGR-VDNQVTPAANQPAANVANQDV--ANPATDAGAL-----NRQSAAD 132
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 138 NGKEVKRQLVNDTKEGAAGFNNDVYKVPAINSSMSGFQGIITLPVTVKNEVQLVHR 197
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 133 TSTDGKA-----VP-QTSDQPGH-----LETVDGKTYVD-ANGQRLKNYS 171
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 198 FSNVDKTSNGYVDFWSELMPVKDSFKGNGPLKQFGLQTINGQQYYIDPTTGQPRKNFL 257
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 172 MVIDGKTYVFDGQTGEAQTDLPKTQANQDNVPSYQANNOAYSNEASSFETVDNYLTAD 231
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 258 LOSGNNWTFDSDTGVGTNALELOFAKGTVSSNEQYRNGNAYSYDDKSIENVNGYLTAD 317
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 232 SWYRPRKILKNGSQWQASSEGDLRPIILMTWPDAAATKAAAYANFWAKEG-LISGSYRQNSA 290
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 318 TWYRPAQILKGTWTTDSKETDWRPILMVWMTLTLQAYLYNMKQHGHLPSALPFFNA 377
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 291 NIDAAQON-----IQSAIEKKIASEGNTNWLDRKMSQFVKSONQWSIASENTVTPNQDH 345
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 378 DADPAELNHYSEIVOQIEKRISETGNTDLRLMHDFTVNNPMWKNKSENVSFGIQ-- 435
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 346 MOGGALLFNSKDTTHANSDFLLNRPFTQTKQKFTTYNAGVELLLANDVDNSPVV 405
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 436 FOGGFLKYNSDLTPYANSDYLLGRMPI-----NIKQTYRGQFLLANDIDNSPVV 489
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 406 QAEQLNHLHYLNMWGDIVMGDKDANFDGVRDAVNANVADLLQIQDYKAKYGTDQNEK 465
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 490 QAEQLNWLYYLLNFGTITANNQANFDSVRVADPNIDADLNIAQDYFNAYGMD-SDA 548
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 466 NAIDLHLSLEAWGNDNVYQDNQNFSLIDNDQSRGMLKAFGYSAYRGNLSNLTATAGL 525
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 549 VSNKHINILEDNWHADPEYFNKIGNPQLTMDDTIKNSLNHGLSDAT-NRWGLDAIVHQS 607
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 526 KNRSANPDSDPV-PNVVFTRAHDSEVQTIKIREKLGKTNADGLTNLTDLNKAFDI 584
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 608 ADRENNSTENVIPNYSFVRAHDNNSQDQIQAIRADVTK-----DYHTFTFEDEQKGIDA 663
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 585 YNQDMNATDKVYYPNNLPMAWMLQNKDVTVRVYVYGDVMTDNGQVMAKTPFYNAIETL 644
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 664 YIQDQNSTVKKNLYNIPASYAILTNKDTIPRVYVYGLDYGQVMEHQTRYDTLTNL 723
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 645 LKGRKYVAGQAVSYKQDWSSG----ILTSVRYGKGANSASDAGNTRNSGMLLNN 700
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 724 LKSRVYVAGGQSM--QTMVGGNNILTSVRYGKGAWTATDTGTDETRTQIGVYVSN 780
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 701 RNPFRAYN--LTLNMGAAHQAQVRPLLSTKDGIAITYLNDSDVDSRQYKTYTDSQGNLS 758
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 781 TPNLKLGVNDKVVLHMGAAHKNQYRAAVLTTTGDVINYTSD---QGAPVAMTDENGDLY 837
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 759 FSASEL-----QSVANAQVSGMIOQVVPVGAADNDVRTSGPSTQATKQGNLYH 806
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 838 LSSHNLVNGKEADTAQGYANPDVSGVLAWVPVPGASDNDQARTAFSTEKNSAYR 897
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 807 QSDALDSQVITYEGFNGFAQSPQYTNNAVIAKNGDLFKSWGITQFEMAPQYVSESDGT 866
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 898 TNAAFDSNVIFPAFNFVYPTTKESERANVRIAQADFFASLGFTSFEMAPQYNSSKORT 957
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 867 FLDSVILNGYAPSDRYDLAMSKNKYKSGQDLANAIKGLQSAKIKVLSDLPVNLNPLG 926
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 958 FLDSIDNGYAPTDRYDLGMSPNKYGTDEDLRNAIQALHKAGLQVMAWDPDQIYNLPG 1017
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 12
US-09-995-749A-2
; Sequence 2, Application US/09995749A
; Patent No. 6867026
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAQUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: BO43388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1781
; TYPE: PRT
; ORGANISM: Lactobacillus reuteri
US-09-995-749A-2

Query Match      30.3%; Score 2191.5; DB 2; Length 1781;
Best Local Similarity 40.9%; Pred. No. 8e-147;
Matches 518; Conservative 170; Mismatches 436; Indels 141; Gaps 34;

QY 36 AADTNNNDGTSV--QVKN-MVPSDPKFDQAQNG--OLAQAMFKAANOADQTATSOVSPAT 91
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 581 AADGNSGDNVTVDQFSKNYATTGNGFDYVKVNGNQVFSGWHATNQSNDKDSQWIIIVLV 640
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 92 DGR-VDNQVTPAANQPAANVANQDV--ANPATDAGAL-----NRQSAAD 132
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 641 NGKEVKRQLVNDTKEGAAGFNNDVYKVPAINSSMSGFQGIITLPVTVKNEVQLVHR 700
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 133 TSTDGKA-----VP-QTSDQPGH-----LETVDGKTYVD-ANGQRLKNYS 171
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 701 FSNVDKTSNGYVDFWSELMPVKDSFKGNGPLKQFGLQTINGQQYYIDPTTGQPRKNFL 760
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 172 MVIDGKTYVFDGQTGEAQTDLPKTQANQDNVPSYQANNOAYSNEASSFETVDNYLTAD 231
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 761 LOSGNNWTFDSDTGVGTNALELOFAKGTVSSNEQYRNGNAYSYDDKSIENVNGYLTAD 820
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 232 SWYRPRKILKNGSQWQASSEGDLRPIILMTWPDAAATKAAAYANFWAKEG-LISGSYRQNSA 290
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
Db      821  TWRPKQILKDGTTWTDKETDMRPILMVWWTNTLTQAYLLNMYKHQGNLLPSALPFFNA 880
QY      291  NIDATON-----ISAEIKKIASEGNTWLRDXMSQFVKSONQWSIASENETVVPNQDH 345
Db      881  DADPAELNHYSEIVQONIEKRISETGNTDRLTLMHDFVTNNPWNKDSNFVSGIQ-- 938
QY      346  MOGGALLPNSKDETHANSDFLLNRPFTQKQKPYFTNYAGYVELLLANDVDSNPV 405
Db      939  FQGGELKYENSULTPYANSDYLLGRMPI-----NIKQTYRGQEPFLANDIDNSPVV 992
QY      406  QAEQLNHLHYLNMWGDIVMGDKDANFDGVRDAVDNVDNADLQIQORDYYKAKYGTQNEK 465
Db      993  QAEQLNWLYYLLNFGTITANNQANFDSVRVADPNIDADLNIAQDYFNAAYGND-SDA 1051
QY      466  NAIDLHLSLEAWSGNDNVKQDNFSLIDNDQSRGMLKAFGYSAYVRGNLSNATLQ 525
Db      1052 VSNKHINILEDWNHDEPFYFNKIGNPQLTMDDTIKNSLNHGLSDAT-NRWGLDAIVHQS 1110
QY      526  KRSANPDSDPV-PNVVFTIRAHDSVQTRIAKIREKLGKTWADGLTNLTLDLNAKADI 584
Db      1111  ADRENNSTENVIPNYSFVRAHDNNSQDQIQAIRHDVTGK----DYHTTFDEQKGIDA 1166
QY      585  NYQDMNATDKVYPNNLPMAWMLQNKDTVRVYGDMYTONGOVMAKTFFYNAIETL 644
Db      1167  YIQDNSTVKKNLYNIPASYAILTNKDTIPRVYGDLYTDGGQYMEHQTRYDTLNL 1226
QY      645  LKGRKYVAGGQVSVYKQDWSG-----ILTSVRYGKGANSASDAGNETFRNSGMLINN 700
Db      1227  LKSRVYVAGGQSM---QTMVGGNNILTSVRYGKGAMTATDTGTDETRTQIGVWVSN 1283
QY      701  RNFPRAYRN--LTLNNGAAKQAYRPLLSYKDGATYLNDSVDSDROYKYTDSQGNLS 758
Db      1284  TPNLKLGVNDKVVLHMGAAHKNQYRAAVLTITTDGVINYTSD---QGAPEVMTDENGDI 1340
QY      759  FSASEL-----QSVANAQVSGMIQVWPVPGAADNODVRTSPSTQATKDGNIYH 806
Db      1341  LSHNLVNGKEADTAVQYANPDVSGYLAWVPVPGASDNQDARTAPSTEKNSGNSAYR 1400
QY      807  QSDALDSQVYIEFGSNFOAQSPOQYTNNAVIAKNGDLFPKSWGITQFEMAPQVYSEBGT 866
Db      1401  TNAAFDSNVIFEAFSNFVYPTTKESERANVRIAQNADFPASLGFTSFEMAPQVNSKORT 1460
QY      867  FLDSVLINGAYESDRYDLAMSKNNKYQSKODLANAIGKLOSAGIKVLSLDPNOLYNLPG 926
Db      1461  FLDSITIDNGYAFTRDYDLGMSBPNKYGTDEDLRNAIQALHKGALQVMADVPDQIYNLPG 1520
QY      927  KEVVTATRVNOYGOAKSGATINKTPVANTRSYGDYQEQYGGKFLDQLKLYPRLFSTKQ 986
Db      1521  KEVATVTRVDRGNVWKDAIINNLYVNTIGGGEYQKYGGAFLDKLQLYEFTKKQ 1580
QY      987  ISTGKPIDPSVKITNWSAKYFNGSNILGRGAKYVL--SEGNKYLN--ADGKLFPTVLNNT 1044
Db      1581  VSTGVAIDPSQKITEWSAKYFNGSNILHRSGLVGLKADGGQYVNLGTTTKQFLPIQLT-- 1638
QY      1045  YQOPQVSANGFISKNGIHYL--DKNGQEVKNRFEKIS--GSWYTFPSDGHQWATKTKIGND 1102
Db      1639  -GEKKQNGEGFVKGNDGNYFYFDLAGNMVKNFTFIEDSVGNWYFFPDQDGMVNGKHFDVD 1697
QY      1103  TYLFPNGKQKGEVWYDGKK--AYYYDDNGRTWTKNGFVEFRVGDGDKWRYFNGDGCTIAI 1161
Db      1698  SY-----GEKGYTFPLKNG-----VSFRG----- 1716
QY      1162  GLVSLDNRTLYFDAYGVYKGVQTVTINGKSYTFDADQGLVQTDNANPA--FQOAGWKL 1219
Db      1717  GLVQTDNGTYFYFDNYGKVRNQTINAGAMIYTLN--ENGLKIKASYNSDAEYPTSTDVGM 1775
QY      1220  LGDNQ 1224
Db      1776  LDQNK 1760
```

RESULT 13

US-09-604-957-4

```
; Sequence 4, Application US/09604957
; Patent No.: 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; CURRENT APPLICATION NUMBER: US/09/604,957
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin ver. 2.1
; SEQ ID NO 4
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Streptococcus mutans
; US-09-604-957-4
```

Query Match 22.3%; Score 1615.5; DB 2; Length 545;
Best Local Similarity 58.9%; Pred. No. 1.4e-106;
Matches 323; Conservative 75; Mismatches 127; Indels 23; Gaps 9;

```
QY      393  LLANDVDSNPVQAEQLNHLHYLNMWGDIVMGDKDANFDGVRDAVDNVDNADLQIQRD 452
Db      1  LLANDIDNSNPVQAEQLNHLHYLNMWGDIVMGDKDANFDGVRDAVDNVDNADLQIASD 60
QY      453  YVYKAYGTDONEKNAIDHLSLEAWSGNDNDVYKQDNFSLIDNDQSRGMLKAF----- 507
Db      61  YLKAHYGVDSKSEKNAINHLSLEAWSDNDPQYNKDTKGAQLPIDNKLRLSLYALTRPLE 120
QY      508  ---GYASVYRGNLSNLATAGLKNRSA--NPDSDPVNPVYFIRAHDSVQTRIAKIREK 563
Db      121  KDSANKNEIRSGLEPVITNSLNNRSAEGKNSERMANYIFIRAHDSVQTRIAKIIQAI- 179
QY      564  KTNADGLTNLTLDLNAKAFDIYQDNMATDKVYPNNLPMAWMLQNKDTVRVYGD 623
Db      180  NPKTDGLT-FTLDELKQAFKIYNEDMRQAKKQYQTSNPTAYALMLSNKSITRLYYGDM 238
QY      624  YTDNGQYMATKTPFYNAJETLLKGRKYVAGGQ--AVSYKQ-----DWS-SGILTSVRY 674
Db      239  YSDDQYMATKSPYDAIDTLTKARIKAYAGGQDMKIYVEGDKSHMDWDYTGVLTSVRY 298
QY      675  GKGANSASDAGNETFRNSGMLINNRPFRAYRN--LTLNNGAAHKSQAYRPLLLSTKD 732
Db      299  GTGANEATDQGESEATKTQGMVITSNPSPSLKLNQNDKVIIVNNGAAHKNQOYRPLLLTTKD 358
QY      733  GIATYLNDSVDSDROYKYTDSQGNLSFSASELQSVANAQVSGMIQVWPVPGAADNODVRT 792
Db      359  GLTSYTSDAAKS--LYRKTNDKGLVFPDASDIQGYLNPQVSGYLAWVPVPGASDQDVRV 417
QY      793  SPSTQATKDGNIYHOSDALDSQVYIEGFSNFOAQSPOQYTNNAVIAKNGDLFPKSWGITQ 852
Db      418  AASNKANATQGVYESSSALDSQLIIEGFSNFOQFTKQSDYTNKIKIAQNVQLFKSWGITS 477
QY      853  FEMAPQVYSSDGTFLDSVILNGYAFSDRYDLAMSKNNKYQSKODLANAIGKLOSAGIKV 912
Db      478  FEMAPQVYSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSGQDQMINAVKALHKSQIQV 537
QY      913  LSDLVNQ 920
Db      538  TADWVPDQ 545
```

RESULT 14

US-09-995-749A-10

; Sequence 10, Application US/09995749A

; Patent No. 6867026

; GENERAL INFORMATION:

; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA

; APPLICANT: DIJKHUIZEN, LUBBERT

```
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: BO43388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Streptococcus mutans
; ORGANISM: Leuconostoc mesenteroides
US-09-995-749A-10

Query Match      22.1%; Score 1597.5; DB 2; Length 545;
Best Local Similarity 58.8%; Pred. No. 2.7e-105;
Matches 323; Conservative 76; Mismatches 125; Indels 25; Gaps 11;

QY      393 LLANDVNSNPVVOAQLNHLHLMWGDIVMGDKDANFDGVRVDAVDVNNADLLQIQRD 452
Db      1 LLANDVNSNPVVOAQLNHLHLMWGDIVMGDKDANFDGVRVDAVDVNNADLLQIQAD 60
QY      453 YYKAKYGTONEKNAIDHLSILEAWSGNDNDYVKDQNNFSLSDNDORSGLKAFV 512
Db      61 YKFLAYGVQNDATANQHLSILEDMSHNDPLYYTDQSGNQLTMDYVHTQLIWSLTSSD 120
QY      513 YRGNLSNLTATAGLKNRS-ANPDSDPVPNVYFIRAHDSVQVTRIARIKLGKTNADGLT 571
Db      121 IKGTMQRFVDYVWDRSNDSTENEALPNYSFVRAHDSVQVTVIAQVSDY--PDVNSL 178
QY      572 NLTDDLKNAFDIYNQDMNATDKVYPNNLPMAYAWMLQNKOTVTRVYVYGDYTDNGQYM 631
Db      179 APTTEQLAAAFKYNDEKLDKCKYQYNMASAYAMLLTNKDTVPRVYVGLYTDGQYM 238
QY      632 ATKTPFYNAIETLLKRIKYVAGGQAVSYKQDWSGILTSVRYKGANASDAGNTETRN 691
Db      239 ATKSPYDAINTLLKARVQVAGGQSMVD---SNDVLTSVRYKQDAMTASDTGTSETRT 295
QY      692 SGWALLINNRPNFRAY--RNLTLNMGAAHKSQAYRPLLLSTKDGATYLNDSVDVSRQYK 749
Db      296 EGIGVIVSNNAELQLEDGHTVTLHMGAAHKNQAYRALLSTTADGLAY--DTD-ENAPVA 352
QY      750 YTDQGNLSFSASELSQSVANAQVSGMIQVWVPVGAADNQDVRTSPSTQATKGNIIYHSD 809
Db      353 YTDANGDLIFTNESIVGVQNPQVSGYLA VWPVPGAQDQDARTASDTTNTSDKVFHSA 412
QY      810 ALDSQVIYEGFSNFOAQAQSPQYTNNAVIAKNGDLFKSWGITQFEMAPQVVSSEDGTFLD 869
Db      413 ALDSQVIYEGFSNFOAFAFATDSSEYTNVIAQNAQDQFKQWGVTSFQLAPQVRSSTDTSLD 472
QY      870 SYLVNGYAFSDRYDLKMSKNNKYGSKQDLANAIKGIQSAGIKVLSLVPNQ 920
Db      473 SIIQNGYAFSDRYDLOYGTPTKYGTADQURDAIKALHAGSIQAIADWVPDQ 523
```

Search completed: February 11, 2006, 19:46:10
Job time : 37.8957 secs

```
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: BO43388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Streptococcus mutans
; ORGANISM: Leuconostoc mesenteroides
US-09-995-749A-10

Query Match      22.1%; Score 1597.5; DB 2; Length 545;
Best Local Similarity 58.8%; Pred. No. 2.7e-105;
Matches 323; Conservative 76; Mismatches 125; Indels 25; Gaps 11;

QY      393 LLANDVNSNPVVOAQLNHLHLMWGDIVMGDKDANFDGVRVDAVDVNNADLLQIQRD 452
Db      1 LLANDVNSNPVVOAQLNHLHLMWGDIVMGDKDANFDGVRVDAVDVNNADLLQIQAD 60
QY      453 YYKAKYGTONEKNAIDHLSILEAWSGNDNDYVKDQNNFSLSDNDORSGLKAFV 512
Db      61 YKFLAYGVQNDATANQHLSILEDMSHNDPLYYTDQSGNQLTMDYVHTQLIWSLTSSD 120
QY      513 YRGNLSNLTATAGLKNRS-ANPDSDPVPNVYFIRAHDSVQVTRIARIKLGKTNADGLT 571
Db      121 IKGTMQRFVDYVWDRSNDSTENEALPNYSFVRAHDSVQVTVIAQVSDY--PDVNSL 178
QY      572 NLTDDLKNAFDIYNQDMNATDKVYPNNLPMAYAWMLQNKOTVTRVYVYGDYTDNGQYM 631
Db      179 APTTEQLAAAFKYNDEKLDKCKYQYNMASAYAMLLTNKDTVPRVYVGLYTDGQYM 238
QY      632 ATKTPFYNAIETLLKRIKYVAGGQAVSYKQDWSGILTSVRYKGANASDAGNTETRN 691
Db      239 ATKSPYDAINTLLKARVQVAGGQSMVD---SNDVLTSVRYKQDAMTASDTGTSETRT 295
QY      692 SGWALLINNRPNFRAY--RNLTLNMGAAHKSQAYRPLLLSTKDGATYLNDSVDVSRQYK 749
Db      296 EGIGVIVSNNAELQLEDGHTVTLHMGAAHKNQAYRALLSTTADGLAY--DTD-ENAPVA 352
QY      750 YTDQGNLSFSASELSQSVANAQVSGMIQVWVPVGAADNQDVRTSPSTQATKGNIIYHSD 809
Db      353 YTDANGDLIFTNESIVGVQNPQVSGYLA VWPVPGAQDQDARTASDTTNTSDKVFHSA 412
QY      810 ALDSQVIYEGFSNFOAQAQSPQYTNNAVIAKNGDLFKSWGITQFEMAPQVVSSEDGTFLD 869
Db      413 ALDSQVIYEGFSNFOAFAFATDSSEYTNVIAQNAQDQFKQWGVTSFQLAPQVRSSTDTSLD 472
QY      870 SYLVNGYAFSDRYDLKMSKNNKYGSKQDLANAIKGIQSAGIKVLSLVPNQ 920
Db      473 SIIQNGYAFSDRYDLOYGTPTKYGTADQURDAIKALHAGSIQAIADWVPDQ 523
```

RESULT 15
US-09-604-957-5
; Sequence 5, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 20:39:23 ; Search time 122.176 Seconds
(without alignments)
4668.162 Million cell updates/sec

Title: US-10-797-821-39

Perfect score: 7230

Sequence: 1 MEKNLRVYKLVKKQWVAIG.....WKNVDGNWYFNRNGLATRW 1365

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications_AA_Main:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7230	100.0	1365	4	US-10-383-930-39
2	7230	100.0	1365	5	US-10-797-821-39
3	3174	43.9	1430	3	US-09-740-274-6
4	3174	43.9	1430	4	US-10-383-930-36
5	3174	43.9	1430	5	US-10-797-821-36
6	3080.5	42.6	1554	4	US-10-383-930-38
7	3080.5	42.6	1554	5	US-10-797-821-38
8	3031	41.9	1590	4	US-10-383-930-37
9	3031	41.9	1590	5	US-10-797-821-37
10	2945.5	40.7	1475	3	US-09-740-274-2
11	2945.5	40.7	1475	4	US-10-383-930-34
12	2945.5	40.7	1475	5	US-10-797-821-34
13	2893.5	40.0	1375	3	US-09-740-274-4
14	2893.5	40.0	1375	4	US-10-383-930-35
15	2893.5	40.0	1375	5	US-10-797-821-35
16	2804	38.8	1518	4	US-10-383-930-40
17	2804	38.8	1518	5	US-10-797-821-40
18	2680.5	37.1	1497	5	US-10-484-218-18
19	2399	33.2	1595	5	US-10-484-218-20
20	2247.5	31.1	1777	5	US-10-484-218-12
21	2227	30.8	2057	4	US-10-417-280A-2
22	2191.5	30.3	1781	3	US-09-995-749A-2
23	2142.5	29.6	1771	5	US-10-484-218-14
24	2104	29.1	1006	5	US-10-484-218-22
25	1597.5	22.1	545	3	US-09-995-749A-10
26	1361	18.8	522	3	US-09-995-749A-11
27	1203.5	16.6	535	3	US-09-995-749A-13

28	1175.5	16.3	787	5	US-10-484-218-16	Sequence 16, Appl
29	1171	16.2	584	3	US-09-995-749A-12	Sequence 12, Appl
30	1115.5	15.4	525	5	US-10-484-218-23	Sequence 23, Appl
31	554.5	7.7	223	5	US-10-484-218-6	Sequence 6, Appl
32	540	7.5	224	5	US-10-484-218-4	Sequence 4, Appl
33	532.5	7.4	223	5	US-10-484-218-10	Sequence 10, Appl
34	480.5	6.6	221	5	US-10-484-218-8	Sequence 8, Appl
35	469.5	6.5	221	5	US-10-484-218-2	Sequence 2, Appl
36	365.5	5.1	2710	4	US-10-011-366-6	Sequence 6, Appl
37	365.5	5.1	2710	4	US-10-354-774-6	Sequence 6, Appl
38	365.5	5.1	2710	4	US-10-271-012-6	Sequence 6, Appl
39	365.5	5.1	2710	4	US-10-729-122-6	Sequence 6, Appl
40	365.5	5.1	2710	4	US-10-729-039-6	Sequence 6, Appl
41	365.5	5.1	2710	5	US-10-729-527-6	Sequence 6, Appl
42	365.5	5.1	2710	5	US-10-727-898-6	Sequence 6, Appl
43	365.5	5.1	2710	6	US-10-728-696-6	Sequence 6, Appl
44	365.5	5.1	2710	6	US-11-001-241-6	Sequence 6, Appl
45	327.5	4.5	2366	4	US-10-011-366-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1

US-10-383-930-39
; Sequence 39, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 39
; LENGTH: 1365
; TYPE: PRT
; ORGANISM: Streptococcus downei
US-10-383-930-39

Query Match	100.0%	Score 7230;	DB 4;	Length 1365;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 1365;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
Qy	1	MEKNLRVYKLVKKQWVAIGVTTVTL	SLAGGVAAADTNNNDGTSVQVKNVPSDPKFD	60
Db	1	MEKNLRVYKLVKKQWVAIGVTTVTL	SLAGGVAAADTNNNDGTSVQVKNVPSDPKFD	60
Qy	61	AAQNGQLAQAAMFAKAAQADQAT	SQVSPATDGRVDNQVTPAANQPAANQVDPAT	120
Db	61	AAQNGQLAQAAMFAKAAQADQAT	SQVSPATDGRVDNQVTPAANQPAANQVDPAT	120
Qy	121	DAGALNRQSAADTSTDGKAVPQTS	DQPGHLETVDGKTYVVDANGORLKNYSVMDGKTYV	180
Db	121	DAGALNRQSAADTSTDGKAVPQTS	DQPGHLETVDGKTYVVDANGORLKNYSVMDGKTYV	180
Qy	181	FDGQTGAQTDLPKTKGQANQNV	PDSPQANNQVSNSEASSETVDNYLTADSWYPRKIL	240
Db	181	FDGQTGAQTDLPKTKGQANQNV	PDSPQANNQVSNSEASSETVDNYLTADSWYPRKIL	240
Qy	241	KNGQSWQASSEGDLRPLMTW	PDAAATKAAAYANFAWKEGLISGSYRQNSANLDAATQNIQ	300
Db	241	KNGQSWQASSEGDLRPLMTW	PDAAATKAAAYANFAWKEGLISGSYRQNSANLDAATQNIQ	300
Qy	301	SAIEKKIASEGNTWNLDRKMSQ	FVKSQNQWSIASENETVYNQDHMQGALLFNSKQDTE	360
Db	301	SAIEKKIASEGNTWNLDRKMSQ	FVKSQNQWSIASENETVYNQDHMQGALLFNSKQDTE	360


```
QY 361 HANSWRLNLRNPTQTQKQKFTTNYAGYELLANDVDNSNPVQAEQLNHLHYLMWNG 420
Db 361 HANSWRLNLRNPTQTQKQKFTTNYAGYELLANDVDNSNPVQAEQLNHLHYLMWNG 420
QY 421 DIVMGDKANFDGVRDADVNNADLLQIQORDYKAKYGTQDNEKNAIDHLSILEAWSGN 480
Db 421 DIVMGDKANFDGVRDADVNNADLLQIQORDYKAKYGTQDNEKNAIDHLSILEAWSGN 480
QY 481 DNDYVKQDNFSLSDNDQSRGMLKAFGYASAYRGNLSNLATAGLKNRSANPDSPDPVNY 540
Db 481 DNDYVKQDNFSLSDNDQSRGMLKAFGYASAYRGNLSNLATAGLKNRSANPDSPDPVNY 540
QY 541 VFIRAHSEVQTRIAKIIREKLGKTNADGLTNLTLDLNLKAPDIYNQDMNATDKVYPPNN 600
Db 541 VFIRAHSEVQTRIAKIIREKLGKTNADGLTNLTLDLNLKAPDIYNQDMNATDKVYPPNN 600
QY 601 LPMAYAWMLQNKDVTTRYVYGDYMDNGQYMAKTTPFYNAIETLLKGRIKYVAGQAVSY 660
Db 601 LPMAYAWMLQNKDVTTRYVYGDYMDNGQYMAKTTPFYNAIETLLKGRIKYVAGQAVSY 660
QY 661 KQDWSGILTSVRYGKGANSASDAGNTETRNAGMALLNNRPNFRAYRNLTLNMGAAHKS 720
Db 661 KQDWSGILTSVRYGKGANSASDAGNTETRNAGMALLNNRPNFRAYRNLTLNMGAAHKS 720
QY 721 QAYRPLLLSTKDGIAITYLNDSDVDSROYKYTDSQGNLSFSAELOSVAQAQVSGMIQVW 780
Db 721 QAYRPLLLSTKDGIAITYLNDSDVDSROYKYTDSQGNLSFSAELOSVAQAQVSGMIQVW 780
QY 781 PVGAADNODVRTSPSTQATKQNIYHOSDALDSQVIYEGFSNFQAFAPQDQYTNNAVIAK 840
Db 781 PVGAADNODVRTSPSTQATKQNIYHOSDALDSQVIYEGFSNFQAFAPQDQYTNNAVIAK 840
QY 841 NGDLFPKSGITQFEMAPQVVSSEDGTFDLSVTLNGYAFSDRYDLAMSNNKYGSKQDLAN 900
Db 841 NGDLFPKSGITQFEMAPQVVSSEDGTFDLSVTLNGYAFSDRYDLAMSNNKYGSKQDLAN 900
QY 901 AIKGLQSAQIKVLSLQVNLQNLVNLPGKEVVTATRVNQYQAKSGATINKTPVANTRYG 960
Db 901 AIKGLQSAQIKVLSLQVNLQNLVNLPGKEVVTATRVNQYQAKSGATINKTPVANTRYG 960
QY 961 DYQEQYGGKFLDQKLPRLPSTQKISTGKPIDPSVKITNWSAKYFNGSNILGRGAKV 1020
Db 961 DYQEQYGGKFLDQKLPRLPSTQKISTGKPIDPSVKITNWSAKYFNGSNILGRGAKV 1020
QY 1021 LSEGNKYLNLADGKFLPTVLNNTYGPQVSANGFIKNGGGHYLDKNGQEVNRFKEIS 1080
Db 1021 LSEGNKYLNLADGKFLPTVLNNTYGPQVSANGFIKNGGGHYLDKNGQEVNRFKEIS 1080
QY 1081 GSWYTFDSGQKATGKTKIGNDTYLFMPNGKQKQKGVYDQKAYYYDDNGRTWTNKGKV 1140
Db 1081 GSWYTFDSGQKATGKTKIGNDTYLFMPNGKQKQKGVYDQKAYYYDDNGRTWTNKGKV 1140
QY 1141 EFRVDGQKWRVFNCGDGTIAIGLVSLDNRTLYFDAYGYQVKGQVTVINGKSYTFDADQGD 1200
Db 1141 EFRVDGQKWRVFNCGDGTIAIGLVSLDNRTLYFDAYGYQVKGQVTVINGKSYTFDADQGD 1200
QY 1201 LVQTDNANPAPQAGWKLGDQNGYRKDGQGLLTGEQITDQKVFQDNGVQVKGGTAT 1260
Db 1201 LVQTDNANPAPQAGWKLGDQNGYRKDGQGLLTGEQITDQKVFQDNGVQVKGGTAT 1260
QY 1261 DASGVLRFYDRDQHQVKGWYSTSDNNWYVNESGQVLTGLQITDQTVFPDDKGIQAK 1320
Db 1261 DASGVLRFYDRDQHQVKGWYSTSDNNWYVNESGQVLTGLQITDQTVFPDDKGIQAK 1320
QY 1321 GKAVWDEGNLRYFPADSGNMLRDRKXVDGNNWYFNENGLATRW 1365
Db 1321 GKAVWDEGNLRYFPADSGNMLRDRKXVDGNNWYFNENGLATRW 1365
```

RESULT 2

US-10-797-821-39

; Sequence 39, Application US/10797821

```
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 39
; LENGTH: 1365
; TYPE: PRT
; ORGANISM: Streptococcus downei
US-10-797-821-39
```

```
Query Match 100.0%; Score 7230; DB 5; Length 1365;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MEKNLYKHLHKQWVAIGVTTVTLSPLAGQVVAADTNNNDGTSVQVKNMVPSPDKPFD 60
Db 1 MEKNLYKHLHKQWVAIGVTTVTLSPLAGQVVAADTNNNDGTSVQVKNMVPSPDKPFD 60
QY 61 AOQNGQLAQAAMFKAANOADQATATSOVSPATDGRVDNQTTPAANOPAANVANODVANPAT 120
Db 61 AOQNGQLAQAAMFKAANOADQATATSOVSPATDGRVDNQTTPAANOPAANVANODVANPAT 120
QY 121 DAGALNRQSAADTSDGKAVPQTSQDPGHLETVDGKTYVVDANGORLKNYSWVIDGKTYV 180
Db 121 DAGALNRQSAADTSDGKAVPQTSQDPGHLETVDGKTYVVDANGORLKNYSWVIDGKTYV 180
QY 181 FDGQTGEAQTDLPKTGQANQDNVPDPSYOANNOAYSNEASSFETVDNYLTADSWYPRKIL 240
Db 181 FDGQTGEAQTDLPKTGQANQDNVPDPSYOANNOAYSNEASSFETVDNYLTADSWYPRKIL 240
QY 241 KNGOSWQASSEGDLRPIILMTWMPDAATKAAVANFWAKEGLISGSYRQNSANLDAATQNIQ 300
Db 241 KNGOSWQASSEGDLRPIILMTWMPDAATKAAVANFWAKEGLISGSYRQNSANLDAATQNIQ 300
QY 301 SAIEKKIASEGNTWLRDKMSQFVKSQNWSIASNETVYPNQDHMQGALLFNSNSKDTE 360
Db 301 SAIEKKIASEGNTWLRDKMSQFVKSQNWSIASNETVYPNQDHMQGALLFNSNSKDTE 360
QY 361 HANSWRLNLRNPTQTQKQKFTTNYAGYELLANDVDNSNPVQAEQLNHLHYLMWNG 420
Db 361 HANSWRLNLRNPTQTQKQKFTTNYAGYELLANDVDNSNPVQAEQLNHLHYLMWNG 420
QY 421 DIVMGDKANFDGVRDADVNNADLLQIQORDYKAKYGTQDNEKNAIDHLSILEAWSGN 480
Db 421 DIVMGDKANFDGVRDADVNNADLLQIQORDYKAKYGTQDNEKNAIDHLSILEAWSGN 480
QY 481 DNDYVKQDNFSLSDNDQSRGMLKAFGYASAYRGNLSNLATAGLKNRSANPDSPDPVNY 540
Db 481 DNDYVKQDNFSLSDNDQSRGMLKAFGYASAYRGNLSNLATAGLKNRSANPDSPDPVNY 540
QY 541 VFIRAHSEVQTRIAKIIREKLGKTNADGLTNLTLDLNLKAPDIYNQDMNATDKVYPPNN 600
Db 541 VFIRAHSEVQTRIAKIIREKLGKTNADGLTNLTLDLNLKAPDIYNQDMNATDKVYPPNN 600
QY 601 LPMAYAWMLQNKDVTTRYVYGDYMDNGQYMAKTTPFYNAIETLLKGRIKYVAGQAVSY 660
```

Db 601 LPMAYAWLONKDTVRVYIGDMYTNGQTMATKTFYNAIETLLKGRKYVAGQAVSY 660
Qy 661 KQDWSSGILTSVRYGKANSASDAGTETNSGMALLINRPNFRAYRNLTLNMGAAHKS 720
Db 661 KQDWSSGILTSVRYGKANSASDAGTETNSGMALLINRPNFRAYRNLTLNMGAAHKS 720
Qy 721 QAYRPLLLSTKGIATYLNDSVDSDRYKYKTSQGNLSPSASELSQVANAQVSGMTQVWV 780
Db 721 QAYRPLLLSTKGIATYLNDSVDSDRYKYKTSQGNLSPSASELSQVANAQVSGMTQVWV 780
Qy 781 PYGAABNDQVRTPSPQATKQGNHYHOSDALDSQVIEGFSNFQAFQSPDOVTNAVIK 840
Db 781 PYGAABNDQVRTPSPQATKQGNHYHOSDALDSQVIEGFSNFQAFQSPDOVTNAVIK 840
Qy 841 NGDLFKSWGITQFEMAPQVSSSEDTFLDSVILNGYAFSDRYDLAMSNNKYGSKQDLAN 900
Db 841 NGDLFKSWGITQFEMAPQVSSSEDTFLDSVILNGYAFSDRYDLAMSNNKYGSKQDLAN 900
Qy 901 AIKGLQASAGIKVLSLQNLNLPQKEVVTATRVNQYQOAKSGATINKTPVANTRSYG 960
Db 901 AIKGLQASAGIKVLSLQNLNLPQKEVVTATRVNQYQOAKSGATINKTPVANTRSYG 960
Qy 961 DYQEOYGGKFLDLOKLYPRLFSTKQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKYV 1020
Db 961 DYQEOYGGKFLDLOKLYPRLFSTKQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKYV 1020
Qy 1021 LSEGNKYNLADGKFLPFLVNLNTYQPOVSANGFISKNGGIHYLDKNGQEVNRFKELS 1080
Db 1021 LSEGNKYNLADGKFLPFLVNLNTYQPOVSANGFISKNGGIHYLDKNGQEVNRFKELS 1080
Qy 1081 GSWYFSDSGKMATGKTIGNDTYLFMPNGKQKEGVWYDGGKAYYDDNGRTWTKGPFV 1140
Db 1081 GSWYFSDSGKMATGKTIGNDTYLFMPNGKQKEGVWYDGGKAYYDDNGRTWTKGPFV 1140
Qy 1141 EFRVDGQKWRVFNCGDTIAGLSLDNRTLYFDAYGYQVKQOTVTTNGKSTFFDADQGD 1200
Db 1141 EFRVDGQKWRVFNCGDTIAGLSLDNRTLYFDAYGYQVKQOTVTTNGKSTFFDADQGD 1200
Qy 1201 LVQTDNANPAPQOAGWKLGNQGYRKDGQLLTGEOTIDGQKVFQDNGVQVKGGTAT 1260
Db 1201 LVQTDNANPAPQOAGWKLGNQGYRKDGQLLTGEOTIDGQKVFQDNGVQVKGGTAT 1260
Qy 1261 DASGVLRFYDRDQGHQVKGWYSTSDNNWVYNESGQVLTGLQTDIGQTVYFDDKGIQAK 1320
Db 1261 DASGVLRFYDRDQGHQVKGWYSTSDNNWVYNESGQVLTGLQTDIGQTVYFDDKGIQAK 1320
Qy 1321 GKAVWDENGLRYFPDADSGNMLRDRWKNVDGNWYFNRNGLATRW 1365
Db 1321 GKAVWDENGLRYFPDADSGNMLRDRWKNVDGNWYFNRNGLATRW 1365

RESULT 3
US-09-740-274-6
; Sequence 6, Application US/09740274
; Patent No. US2002031826A1
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172

; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-6

Query Match 43.9%; Score 3174; DB 3; Length 1430;
Best Local Similarity 47.3%; Pred. No. 1.7e-182;
Matches 677; Conservative 213; Mismatches 436; Indels 104; Gaps 38;

Qy 1 MEKNLRYKLHKYKQWVAIGVTT--VTLSFLAGGQVVAADTNNNDGTSV-----QVNMV 53
Db 1 METKRYKMKHKYKQWVAVASGLITITGTTILGSSVSAETEOQTSKDVQVTKSEDDKAA 60

Qy 54 PSDPKFDA---QAQNGQLAQMFKAAQADOTATISQVSPATDGRVDNOVTPAANQPAAN 109
Db 61 SESSQTDAPKTKQAQTEQ--TQAQ--SQANVAD--TSIS-----ITKETPSQNIITTOANSDDKT 113

Qy 110 VANQDVANPATDAGALNRQSAADTSTDGKAVP-----QTSQDPG-----H 149
Db 114 VNTTKSEEAQTSSEBTKESEEAQTASSQALTKQAKAELTKQRTAAQENKNPVDLAAIPN 173

Qy 150 LETVVGKTYVVDANGORLKNYSWVIDGKTYYPDGOTGEGAQTDLP-----KTQOANQDNVPD 205
Db 174 VQIIDGKTYIIGSDQPKPKNFALTVNNKVLVYDPKNTG--ALTDTSQVQFKOGLTKLAN-- 229

Qy 206 SYQANNOAYSNEASFETVDNYLTADSWYRPRKILKNGQSQWASSEGLRPLIMTWMPDA 265
Db 230 DYTTHQIVNFENTSLTIDNVTADSWYRPRKILKNGKTTWATSESDLRPLMLSMWPK 289

Qy 266 ATKAAVANFWAKEGLISG--SYRQNSA--NLDAATQNIQSAIEKKIASEGNTNWLKMSQ 322
Db 290 QTQIAYLNMNQOGLGTGENTYADSSQSLNLAQTVQVKIETKISQTOQTQWLRLDIINS 349

Qy 323 FVKSONQWISIASENEFVTPNODHMOGGALLFSNKSDEHANSDRLLNRPITFOTCKQY 382
Db 350 FVKTPQNNWSQTESDTSAGEKDLQGGALLYSNDSKTAAYANSYDRLLNRPITFOTCKPY 409

Qy 383 FTTNYA--GYELLANDVNSPNVQAEQLNLHLHYLMNWGDIVMGDKDANFDGVRVDAVN 441
Db 410 FEDNSSGGVDFLLANDIDNSNFPVQAEQLNLHLHYLMNYGSIIVANDEANFDGVRVDAVN 469

Qy 442 VNADLLQIORDYYKAKYGTQDQNEKVAIDHLSLLEAWSGNDNDYVKDQNNFSLSDNDQRS 501
Db 470 VNADLLQIASDYLKAHYGVDKSEKNAIHNLSLLEAWSNDNDPOYNKDTKGAQLPIDNKLRL 529

Qy 502 GMLKAF-----CYASAYRCNLSNLATAGLKNRSA--NPDSDPVPNTYVFIABHSEVOT 552
Db 530 SLLYALTRPLEKADASKNEIRSGLEPVTITNSLNNRSABGKNSERMANYIFIRABHSEVOT 589

Qy 553 RIAKIIREKLGKTNADGLTNLTDLNLKAFDLYNODMNATDKVYYPNNLPMAYAWMLONK 612
Db 590 VIAKIIRKQAI--NPKTDGLT--FTLDELKQAFKIYNEDMEQAKKYTQSNIPYALMLSNK 647

Qy 613 DTVTRVYVYDMDYDNGQYMATKTPFYNAIETLLKGRIKYVAGGQ--AVSYKQ-----DW 664
Db 648 DSITRLYYGDMYSDDDQYMATKSPYDAITDLLKARIKYAAGGQDMKITVYVEGDKSHMDW 707

Qy 665 S-SGLITSVRVYKGANASDAGNTETNRSGMALLINRPNFRAYRN--LTLNMGAAHKSQ 721
Db 708 DYTGLVTSVRVYGTGANEATDQGEATKQGMNAVITSNPNLSKLNQNDKVINMGAAHKNQ 767

Qy 722 AYRPLLLSTKGIATYLNDSVDSDRYKYKTSQGNLSPSASELSQVANAQVSGMTQVWV 781
Db 768 EYRPLLLSTKGIATYLNDSVDSDRYKYKTSQGNLSPSASELSQVANAQVSGMTQVWV 826

Qy 782 VQAADNODVRTSPSTQATKQGNHYHOSDALDSQVIEGFSNFQAFQSPDOVTNAVIK 841
Db 782 VQAADNODVRTSPSTQATKQGNHYHOSDALDSQVIEGFSNFQAFQSPDOVTNAVIK 841

```
Db 827 VGASNDQVRVAASNKANATQVYESSALDSQLIEGFSNFQDFVTKDSYTNKKIAQN 886
Qy 842 GDLFKSWGITOPEMAPQYVSSSDGTFDLSVILNGYAFSDRYDLAMSKNNKYSGKODLANA 901
Db 887 VQLFKSWGVTSEMAPQYVSSSDGTFDLSIIQNGYAFEDRYDLAMSKNNKYSGQDDMINA 946
Qy 902 IKGLQASAGIKVLSLVPNQLYNLPKGEVVTATRVNQYGOAKSGATINKTPYVANTRSYG- 960
Db 947 VKALHKSQIADVPDQIYNLPKGEVVTATRVNDYGEYRSDSEIKNTLYAANTKSNKG 1006
Qy 961 DYQEOYGGKFLDLDLQKLYPRLEFSTQISTGKPIDSPSVKITNWSAKYFNGSNILGRGAKV 1020
Db 1007 DYQAKYGGAFSELAAKYPSIFNRTQISNGKKIDSEKITAWKAKYFNGTNILGRGVY 1066
Qy 1021 LSE--GNKYLNIADGKFLPTVLNNTYGOQVPSANGFISKNGGIHYLDKNGQEVKNRP-K 1077
Db 1067 LKDNASDKYFELKGNQTLPLPKQMTN-----KEASTGFVNDGNGMTFYSTSGYQAKNSFVQ 1121
Qy 1078 EISGSWYFSDSGKMATGKTKIGNDTYLPMPNGKQLKEGVW--YDGKAYYYDDNGRTWT 1135
Db 1122 DAKGNWYFDNNGHMYVGLQNGEYQVYFSLNSGVLRESFLENADGSK-NYFCHLGNRYS 1180
Qy 1136 NKGFEFRVGDQKRWYFNGDGTIAGLVSLDNRTLYFDAYGYQVKGQTVT-INGKSYTF 1194
Db 1181 N-GYYSF--DNDSKRWYFDASGVMAVGLXTINGNTQYFDQDGYQVKGAWITGSDGKKYF 1237
Qy 1195 DADQGLVOTDNANPAQOAGWKLLGDNQWY-RKDGQLLTGEQITDQKVFQDNGVQ 1253
Db 1238 DDGSGNMAVRPANDK-----NGDWYLLNSDGLYGVQTINGKTYFYFGDQKG 1286
Qy 1254 VKGQTATDASGVLRFYDRDQHQVGKGYSTSDNNWYVYNGSGVLTGLQITDQGVYFD 1313
Db 1287 IKGKIITD--NGKLKYFLANSGLARNIFATDSQNNWYFPGSDGVAVTGSQTTAGKKLYPA 1345
Qy 1314 DKGIOAKGKAVDENGRLRYFDADSGNMLDRWK-NVDGNWYFNRNGLA 1362
Db 1346 SDGKQVKGsfv--TYNGKVHYHADSGELQVNRFEADKOGNWTYLLDSNGEA 1394

RESULT 4
US-10-383-930-36
; Sequence 36, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 36
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-36

Query Match 43.9%; Score 3174; DB 4; Length 1430;
Best Local Similarity 47.3%; Pred. No. 1.7e-182;
Matches 677; Conservative 213; Mismatches 436; Indels 104; Gaps 38;

Qy 1 MEKNRLYKLHKVKKQWALGVTT--VTLSPFAGGQVVAADTNNDGTSV-----QVNRKV 53
Db 1 METKRYKMHKVKGVHVTVAASGLITLGTTLTGLSSVSAETEQSDTKVTKSEDDKAA 60
Qy 54 PSDPKFDA----QAQNGQLAQMFKAAQADOTATSOVSPATDGRVDNVOVTPAAQPAAN 109
Db 61 SESSQTDAPKTKQAQTEQ-TQAQ-SQANVAD-TSIS-----ITKETPSQNIITQANSDDKT 113
```

```
Qy 110 VANQDVANPATDAGALNRQSAADTSTDGKAVP-----QTSDDPG-----H 149
Db 114 VTNTKSEBAQTSEERTKQSEEAQTASSQALTQAKAELTKQRTAAQENKPNVDLAAIPN 173
Qy 150 LETVDGKTYVDANGORLKNYSWVIDGKTYFYFDGOTGEAQTDLP---KTQGANODNVDP 205
Db 174 VKQIDGKYYYIGSDGQPKNFALTVNNKVLFPDKNTG-ALTDTSQYQFKGLTKLNN--- 229
Qy 206 SYQANNQAYSNEASFETVDNYLTADSWYRPRKILKNGSQWASSEGDLRPLIMTWMPDA 265
Db 230 DYTPHQVNFENTSLTIDNYVTADSWYRPRKILKNGKMTWTASSEDLRPLLMSSWMPDK 289
Qy 266 AKKAAAYANFWAKEGLISG-SYRQNSA--NLDAATQNIQSAIEKKITASEGNTNWLDKMSQ 322
Db 290 QTQIAYLYNMQQGLGTGENTADSQBSLMLAAQTQVKIETKISQTOQTQOQMLRDINS 349
Qy 323 FVKSONQMSIASENETVYPNODHMOGGALLFNSKSDTEHANSMDRLLNRPFTQTGKQKY 382
Db 350 FVKTOPNWSQTESDTSAGEKDLQGGALLYNSDKTAYANSDYLLNRTPTFSQTGKPKY 409
Qy 383 FTTNYA-GYELLANDVDNSNPVQAEQNLNHLHYLMNMGDIYVMGDKDANFDGVRVDAVN 441
Db 410 FEDNSSGGYDFLLANDIDNSNPVQAEQNLNHLHYLMNYGSIIVANDPEANFDGVRVDAVN 469
Qy 442 VNADLLQIORDYKAKYGTQNEKVAIDHLSILEAWSGNDNDYVKQNNFSLSIDNDQRS 501
Db 470 VNADLLQIASDYLKAHYGVDSKNAIINHLSILEAWSNDNDPOYNKDTTGAQLPIDNKLRL 529
Qy 502 GMLKAF-----GYASAYRGNLSNLATAGIKNRS--NPDSDPVPNTVYFIRAHSEVOT 552
Db 530 SLLYALTPLPEKADSNKNEIRSGLEPVITNSLNRSAGKNSERMANVYFIRAHSEVOT 589
Qy 553 RIAKIIREKLGKTNADGLTNLTDDLKNAFDIYNODMNAIDKVVYPNNLPMAWMLQNK 612
Db 590 VIAKIIKAQI-NPKTDGLT-FTLDELKQAFKIYNEDEMRQAKKKTQSNIPYALMLSNK 647
Qy 613 DTVTRVYGDVMTDNGQWMAKTPPYNAIETLLKRIKTYVAGGQ--AVSYKQ-----DW 664
Db 648 DSITRLYYGDMYSDDGQWMAKTPPYDAITLLKARIKYAAGGQDMKITYVEGDSKSHMDW 707
Qy 665 S-SGILTSVRYCKGANSADAGNTRNSGALLNRPNPRAYRN--LTLNMGAAHKSQ 721
Db 708 DYTGVLTSVRYGTGANEATDQSEATKQGMVITSNPNPSLKLNDQDKVIVNMGAAHKSQ 767
Qy 722 AYRPLLLSTKDGATYLNDSVDVSRQYKYTDSQGNLSFSASELSQVANAQVSGMIQVWVP 781
Db 768 EYRPLLLTTKDLTSTYSDAAAKS-LYRKTNDKGELVFDASDIQGYLNPQVSGYLAVWVP 826
Qy 782 VQAADNQVTRTSPSTQATKQGNIIHQSDALDSQVIYEGFSNFOAFAQSPDQYTNVIAKN 841
Db 827 VGASNDQVRVAASNKANATQVYESSALDSQLIEGFSNFQDFVTKDSYTNKKIAQN 886
Qy 842 GDLFKSWGITOPEMAPQYVSSSDGTFDLSVILNGYAFSDRYDLAMSKNNKYSGKODLANA 901
Db 887 VOLFKSWGVTSEMAPQYVSSSDGTFDLSIIQNGYAFEDRYDLAMSKNNKYSGQDDMINA 946
Qy 902 IKGLQASAGIKVLSLVPNQLYNLPKGEVVTATRVNQYGOAKSGATINKTPYVANTRSYG- 960
Db 947 VKALHKSQIADVPDQIYNLPKGEVVTATRVNDYGEYRSDSEIKNTLYAANTKSNKG 1006
Qy 961 DYQEOYGGKFLDLDLQKLYPRLEFSTQISTGKPIDSPSVKITNWSAKYFNGSNILGRGAKV 1020
Db 1007 DYQAKYGGAFSELAAKYPSIFNRTQISNGKKIDSEKITAWKAKYFNGTNILGRGVY 1066
Qy 1021 LSE--GNKYLNIADGKFLPTVLNNTYGOQVPSANGFISKNGGIHYLDKNGQEVKNRP-K 1077
Db 1067 LKDNASDKYFELKGNQTLPLPKQMTN-----KEASTGFVNDGNGMTFYSTSGYQAKNSFVQ 1121
Qy 1078 EISGSWYFSDSGKMATGKTKIGNDTYLPMPNGKQLKEGVW--YDGKAYYYDDNGRTWT 1135
Db 1122 DAKGNWYFDNNGHMYVGLQNGEYQVYFSLNSGVLRESFLENADGSK-NYFCHLGNRYS 1180
```

QY 1136 NKGEVFRVDDGQKWRVFNFGDGTIAIGLSLDNRTLYFDAYGVQVKGQVTV-INGKSYTF 1194
Db 1181 N-GYISF--DNDSKWRVFDASGVMAVGLKTINGNTQYFDQDGYQVKGAMITGSGKKRYF 1237
QY 1195 DADGDLVQTDNANPAPQAGWKLJLGDNDQWY-RKDGQLLTGEQITDQKVFQDNGVQ 1253
Db 1238 DGSNGMAVNRFANDK-----NGDWYILNSDGIALVGQTINGKTYTFYFGQDGK 1286
QY 1254 VKGGTATDASGLRFFYDRDQGHQVKGWYSTSDNNWYVNESQVLTGLQITDQGTVFYD 1313
Db 1287 IKGKIITD-NGKLKYLFLANSBELARNIPATDSQNNWYVFGSDGVAVTGSQTIAGKLYFA 1345
QY 1314 DKGIQAGKAVDENGRLRYFDADSGNMLDRWK-NVDGNWYVFNRLGLA 1362
Db 1346 SDGQVKGFSV-TYNGKVHYTHADSGELQVNRFEADKDGWYILDSNGEA 1394

RESULT 5
US-10-797-821-36
; Sequence 36, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 36
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-36

Query Match 43.9%; Score 3174; DB 5; Length 1430;
Best Local Similarity 47.3%; Pred. No. 1.7e-182;
Matches 677; Conservative 213; Mismatches 436; Indels 104; Gaps 38;

QY 1 MEKNRLYKLHKVKKQWVAIGVTV--VTLSFAGGQVVAADTNNDGTSV-----QVKNV 53
Db 1 METKRYKMHKVKKHWVTVAVASGLITLGTTLGSSVAETEQQTSKDVVTKSEDDKAA 60
QY 54 PSDPKFDA---QAQNGLOAAMFKAAANOADOTATSOVSPATDGRVDNVTPAANQPAAN 109
Db 61 SESSQTDAPKTKQAQTEQ-TQAQ-SQANVAD-TSTS-----ITKTPSQNITQANSDDKT 113
QY 110 VANQDVANPATDAGALNRQSAADTSTDGKAVP-----QTSQPG-----H 149
Db 114 VTNKSEEAQTSERTKQSEEAQTASSQALTAQKAEKTKQRTAQENKNPVDLAAIFN 173
QY 150 LETVDGKTYVDANGRLKNSWIDGKTYFDFGQTGEAQTDLP-----KTGAQNOANVPD 205
Db 174 VKQIDGKYYIGSDGQPKNFALTVNNKVLVDFDKNTG-ALTDTSTQYQFKQGLTKLNN--- 229
QY 206 SYQANNOAYSNEASSPETVDNLTADSWYRPRKILKNGOSWASSRGLRPLILMTWPD 265
Db 230 DTPRHQIVNFENTSLTIDNTVTDADSWYRPRKILKNGKTWASSRDLRPLILMTWPD 289
QY 266 ATKAAYANFWAKEGLISG-SYRQNSA--NLDAATQNIQSAIEBKTIASEGNTNWLDRDKMSQ 322

Db 290 QTQIAYLNMNQQGLGTGENYTTADSSQESLNLAAQTQVQKIETKISQTOQTQWLRLDIINS 349
QY 323 FVKSQNWSTASNETVYPNQDHWOGALLFNSKQDEHANSQWRLINRPTQTGQKQY 382
Db 350 FVKTQNMNSQTESDTSAGEKQHLQGALLYSNDSKTAYANSYRLLNPTPTQTGPKPY 409
QY 383 FTTNYA-GYELLILANDVNSNPVQAEQLNHLHYLMNWGDI VMGDKDANFDGVRVDAVDN 441
Db 410 FEDNSSGGYDFLLANDIDNSNPVQAEQLNHLHYLMNYSIVANDPEANFDGVRVDAVDN 469
QY 442 VNADLLQIORDYKAKYGTQDQNEKNAIDHLSILEAWSGNDNDYVKQNNFSLIDNDORS 501
Db 470 VNADLLQIASDYLAHYGVDKSEKNAIHLNLSILEAWSNDNDPQNKOTKGAQLPDKLRL 529
QY 502 GMLKAF-----GYASAVRGNLSNLATAGLNRSA-NPDSDPVNPVIRAHDSVQ 552
Db 530 SLLYALTRPLEKADSNKNEIRSGLEPVTNSLNNRSABGNKSERMANIYIFIRAHDSVQ 589
QY 553 RIAKIIREKLGKTNADGLTNLTLDLNLKAFDIYNQDMNATDKVYYPNNLPMAYAMLQNK 612
Db 590 VIAKIIKAQI-NPKTDGLT-FTLDLQKAFKIYNEDMRQAKKYTOSNIPTALMLSNK 647
QY 613 DTVTRVYGYDMYTDNQYMATKTPFYNAIETLLKGRIKYVAGGQ--AVSYKQ-----DW 664
Db 648 DSITRLYYGDMYSDDCQYMATKSPYYDAIDTLIKARIKYAAGQDMKITYVEGDKSHMDW 707
QY 665 S-SGLITSVRYGKANSASDAGNTETNSGMALLNNRNFRAYRN--LTLNWAHAKSQ 721
Db 708 DYTGLVTSVRYGTGANEATDQSEATKTQGMVITSNNPSLKLNQNDKVINNNGAAHKNQ 767
QY 722 AYRPLLLSTKDGITATYLNDSVDROYKYVTDSONLSFASASELQSVANAOVSGMIQVWP 781
Db 768 EYRPLLLTTKDGILTSYTSDAAKS-LYRKTNDKGLVFDASDIQGLNPNQVSYLAWVP 826
QY 782 VGAADNQDVRTSPSTQATKDGNIYHQSADALDSQVIYEGFSPNFOAQAQSPDQYTNVIAKN 841
Db 827 VGASDNQDVRVAASNKANATQVYESSALDSQLIYEGFSNFDFTKDSYTNKKIAQN 886
QY 842 GDLFKSWGITQFEMAPQYUSSDEGTFLDVILNGYAFSDRYDLAMSKNNKYGSKQDLANA 901
Db 887 VOLFKSWGVTSPFEMAPQYVSSDEGTFLDVILNGYAFSDRYDLAMSKNNKYGSKQDLANA 946
QY 902 IKGLQSAIGIKVLSLDLVPNLPGKEVTVTRVNOYGOAKSGATINKTPYVANTRSYG- 960
Db 947 VKALHKSGLQVIADWVFDQIYNLPGKEVTVTRVNDYGEYRKDSEIKVTLAANTKSNKG 1006
QY 961 DYQEQYGGKFLDLDLQKLYPRLFSTKQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKYV 1020
Db 1007 DYQAKYGAFTSELAAPYSIFNRTQISNGKKIDPSEKITAWKAKYFNGSNILGRGAKYV 1066
QY 1021 LSE--GNKYLNLADGKFLPVTVLNNTYGOQVPSANGFISKNGGIHYLDKNGQVQRFP-K 1077
Db 1067 LKDNASDKYFELKGNQTYLPKQMTN-----KEASTGVNDGNGMTFYSTSGYQAKNSFVQ 1121
QY 1078 ETSQSWYFSDSGKMATGKTIGNDTYLPMNGKOLKEGV--YDCKKAYVYDNGRTWT 1135
Db 1122 DAKGNWYFDDNNGHVMYGLQQLNGEVOYFLSNGVQLRESFLENADGSK-NYFGLHNRYS 1180
QY 1136 NKGFVEFRVDGQKWRVFNFGDGTIAIGLSLDNRTLYFDAYGVQVKGQVTV-INGKSYTF 1194
Db 1181 N-GYISF--DNDSKWRVFDASGVMAVGLKTINGNTQYFDQDGYQVKGAMITGSGKKRYF 1237
QY 1195 DADGDLVQTDNANPAPQAGWKLJLGDNDQWY-RKDGQLLTGEQITDQKVFQDNGVQ 1253
Db 1238 DGSNGMAVNRFANDK-----NGDWYILNSDGIALVGQTINGKTYTFYFGQDGK 1286
QY 1254 VKGGTATDASGLRFFYDRDQGHQVKGWYSTSDNNWYVNESQVLTGLQITDQGTVFYD 1313
Db 1287 IKGKIITD-NGKLKYLFLANSBELARNIPATDSQNNWYVFGSDGVAVTGSQTIAGKLYFA 1345
QY 1314 DKGIQAGKAVDENGRLRYFDADSGNMLDRWK-NVDGNWYVFNRLGLA 1362

Db 1346 SDGQVKGSGFV-TYNGKVHYHADSGELQVNRFEADKGNWYVLDNGEA 1394

RESULT 6
US-10-383-930-38
; Sequence 38, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 38
; LENGTH: 1554
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-383-930-38

Query Match 42.6%; Score 3080.5; DB 4; Length 1554;
Best Local Similarity 43.9%; Pred. No. 8.4e-177;
Matches 667; Conservative 213; Mismatches 446; Indels 193; Gaps 34;

QY 1 MEKNLYKLHKVKKQVAIGVTITLSFLAGGVAA-----DTNNNDG 44
DB 1 MEKLLHKLHKVKKHWTITAVASIGLVSLVAGTVAEDKVANDTTAQTGVDTGDDQA 60

QY 45 TSQVKNVPSDPKFDQAQNGQLAQAMFEKAANQADQ----TATSOVSPATDGRVDNQVT 100
DB 61 TTNDANTNTTDTTDAQSANTNQ-DQAGSDQSNNOQAQDQTANTDNQADNSQTDN--- 116

QY 101 PAANQPAANVANDVANPATDAGALNRQSAADTST----DGKAVPTSDQPG-----HLE 151
DB 117 -----NOATDQATSPATDGTSGVQRDAANVATAADQEGQTAPSEQEKSAALSLDNVK 168

QY 152 TVDGKTYVDANGQRLKNSWIDGKTYTYFDGOTGEAQDTLPKGTQANQDNVPSQANN 211
DB 169 LIDGKTYVQADGSYKKNFAITVNGMLYFSDTGALSTSTYSFSQGTNLNLDVDFSSH 228

QY 212 QAYSNEASSFETVDNLTADSWYRPRKILKNGSQWQASSEGDLRPLMTWPDAAATKAAY 271
DB 229 KAYDSTAKSFELVNGVLTANSWYRPAIGILRNGQTHEASNDLRLPVLMSWPDKDTQVAY 288

QY 272 ANPW-----AKEGLISGSYRQNSANLDAATQNIQSAIEKKIASBGNWNNLWLRDMSQFVKS 326
DB 289 VNYMNYLSANETEVTNETSQVDLNKEA--QSIQTKEQKITSDNSTQMLRTAMEAFVAA 346

QY 327 QNCWSTASENETVYPNQDHMOGALLFNSKQTEHANSWRLNRRPTQTGKQKQFTT- 385
DB 347 QPKWMNSTEN---FNKGDHLQGLLYTNSDLTFPWANSDYRLNRRPTQDQGTCKKYFTG 403

QY 386 NVAGYELLANDVDNPNVYQAEQLNHLHYLMNWGDI VMGDKDANFDGVRVDAVDNVDNAD 445
DB 404 GEGGYEFLLSNDVDNPNVYQAEQLNHLHYLMNWGDI VMGDKDANFDGVRVDAVDNVDNAD 463

QY 446 LIQIQRDYKAKYGTQDNEKNAIDHLSILEAMSGNDNYVKQDNFSLSDINDQSRGMLK 505
DB 464 LLQVSNYFKDNYKVYTDSEANALAHISILEAWSLNDNQYNEDTNGTALSIDNSSRLTSLA 523

QY 506 AFGYASAYCNSLNTATAGLKNRSPNDS--DPVPNYVFIKRAHDSVQTRIAKIREKLG 563
DB 524 VTKQPGQIDLSNLISSESNKERNANDTAYGDTIPTYSFVRAHDSVQTRIAKIREKLI- 582

QY 564 KTNADGLNLTLDLNLKAFDIYNQDMNATDKVYPNNLPMAYAMWLQNKDQTVTRVYVGM 623
DB 583 DTNSDGYT-FTLDQKDAFKIYNEDMAKVNTYTHYNI PAAYALLLSNESVPRVYVGD 641

QY 624 YTDGQYMATKTPFYNAIETLLKGRIKYVAGQAVS-YKQDWSSGILTSVRYCKGANSAS 682
DB 642 YTDGQYMAKSPYDYDAIATMLQGRYAYVSGQSEVHKVNGNQILSSVRYGQDILMSAD 701

QY 683 DAGNTE-TRNSGMALLINRPNF-RAYRNLTLNMGAAHKSQAYRPLLLSTKQGIATYLD 740
DB 702 DTQGTDLSTSLGVLTVSNDPNLGLGDSLTVMNGRAHANQAYRPLILGTGQGVQSYLKD 761

QY 741 SDVDSRQYKTDQGNLSPSASELOSANAOVSQVGMIVQWVPVGAADNQDVRTSPSTQATK 800
DB 762 SDTNI--VKYTDANGNLTFADDIKGYSTVDMSGVLAVWVPVGAQGDQVRVAADNTQKA 819

QY 801 DGNVHQSDALDSOVIYEGFSNFOAFAQSPDOYTNNAVIAKNGDLFKSWGITQFEMAPQVY 860
DB 820 DGSLKTSAAALDSQVIYEGFSNFOFANNDADYTNKIAENADFFKGLGITFEMAPQVY 879

QY 861 SSEDGTFLDSVILNGVAFPSDRYDLAMSKNKYGSQDLANAIKGLQSAQIKVLSLDVPPNQ 920
DB 880 SATDGSFLDSIIQNGYAFSDRYDLAMSKNKYGSQDLANALKALHANGIQAIADWVPQ 939

QY 921 LYNLFQKVVVATRVNOYQAKSGATINKTPYVANTRSYG-DYQEQYGGKFLDDLQKLYP 979
DB 940 IYQLPGEVVVTAKRTNSYGNPTFDVAINNALYATNTKSSGSDYQAYGGAFLDELKAKYP 999

QY 980 RLFPSTKQISTGKPIDPSVKITNWSAKYFNGSHILGRGAKYVLSSEG--NKYLMADGLFL 1037
DB 1000 DMFTVMNISTGKPIDPSTKIKQWEAKYFNGTNVLGKAGYVLSDDATGYFTVNNENGDFL 1059

QY 1038 PTVLANTYQPOVSANGFISKNGGIHYLDKNGQEVKNRPEISGSWYVYFSDSKMATG-- 1095
DB 1060 PA-----SFTGDQNAKTGYFYDGTGMAYYSTSGKAVNSFIYEGHYVYFDKGHWVTGSY 1115

QY 1096 KTKIGNDTYLFMPNGKQKLEGVYDYGK-KAYYYDDNGRTWTNKG-----FVEFRVDGQDK 1149
DB 1116 KAEDGND-YYFLPNGIQMRDAIYQDAQNSYYVGRGTILY--KGDWYVFPVDPNNANKTV 1172

QY 1150 WRYFNGDGTIALGLVSLDNRTLYFDAYGYOVKGQVTTNGKSYTFDADOG----- 1199
DB 1173 FRYFANNNVMAIYGRNMYGQTYFYFDENGFGQAKQLLTDGKGYHFDGNGAMAKNFVNV 1232

QY 1200 --DLVQTDNANPAQOQ----- 1214
DB 1233 GDDWTYMDGNGNAVKGQFPVNNQILYFNPETGVQVKGQFITDAQRTSYSDANSALKKS 1292

QY 1215 -----AGWKLIGDN--QNGY--RKDGQLLTGEQTIDGQKVPFQ-DN 1250
DB 1293 GFTFNGSDWYAEANGYVYKGFQKVAENQDQWYVDPQTTGKQAKGAADVDRDLNFPDS 1352

QY 1251 GVQVKGGTATDASGVLRFYDRDQGHQVKGWYSTDDWVYVNESQVLTGLQITD---- 1306
DB 1353 GVQVKGDFATDESNGTSTFYHGDNGDKVGGFFTTGNNAYYADNNGNLVYKGFQEIDGKY 1412

QY 1307 -----GQTVYFD-DKGIQAKGKAVDENGMLRYFDADSGMMLRDRWKN 1348
DB 1413 HFDEVTGQAKGAALVNGQQLYFDVDSGLQVKGDFVTDQCGNTSYVDVNSGD-----KK 1466

QY 1349 VDG-----NWYFNRNG 1360
DB 1467 VNGFPTTGDNWAYADGQG 1485

RESULT 7
US-10-797-821-38
; Sequence 38, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09


```
; SEQ ID NO 37
; LENGTH: 1590
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-383-930-37

Query Match          41.9%; Score 3031; DB 4; Length 1590;
Best Local Similarity 41.8%; Pred. No. 8.4e-174;
Matches 672; Conservative 217; Mismatches 436; Indels 292; Gaps 36;

QY 1 MEKNRYKLHKVKKOWAVGVTTVT--LSFLAGQVVAADTNNNDGTSVQVNNKVPSPDKPF 59
DQ 1 MEKNVFPKHVKKRWVTLVSASATMLASALGASVASAUTDITASDDSNQA--VITGD--- 55
QY 60 DAQAQNGQLAQAAMFAKAAQADOTATTSQVSPAIDGRVDNQVTPAANQPAANVANQDV--- 115
DQ 56 --QTTNQATDQTSIAA-----TATSEQASASTDAATDQ--ASAAEQTGCTASTDAAQT 106
QY 116 ---ANPA-----TDAGALNRQSAADTSTDGKAVPQTSDDPGHLETVDGKTYV 160
DQ 107 TTANAEKAVPTENENQGTDEMLAEAKNVATAESD--SIPDLAKKSNVQVQDGYKYY 164
QY 161 DANGQRLKKNYSMVIDCKTYVFGQTG---EAQTDLPKTGOANQDNVPDSYOANNOAYSN 216
DQ 165 DQDGNVKNFVSVGDKIYYFD-ETGAYKDTSKVDADKSSAVSQA-TIFAANPEAYST 222
QY 217 EASSPETVNYLTADSWYRPRKILKNGQSQWASSEGDLRPLMTWMPDAATKAAAYANFWA 276
DQ 223 SAKNFEAVDNYLTADSWYRPSKILKDKTWTESGDDPRPLMAMWPDTETKRNYVYVNW 282
QY 277 KEGLISGSY--RONSANLDAQNTIOSAEKKTASEGNTNWLDRKMSQVFKSONWSIAS 334
DQ 283 KVVGIDKTYTAETSQADLTAAAEVLQARIEQKITSNNTKWLREATSAFVKTPQPMNGES 342
QY 335 ENETVPNODHMOGGALLFSNKKD--TEHANSWRLNLRNPTFOTGK--QKYFTTN----- 388
DQ 343 EK-----PYDHLQNGALLFDNQTDLTPTDQSNVRLNLTPTNQTSGLDSRFTYNPNDPLG 398
QY 389 GYELLANDVNSNPVQAEQLNHLHYLMNMWDIVMGDKDANPQVGRVDAVNDAVNADLLQ 448
DQ 399 GYDFLLANDVNSNPVQAEQLNHLHYLMNFGSIYANDADANFDSIRVDAVNDAVNADLLQ 458
QY 449 IQRDYVYKAKYGTQDNEKNAIDHLSILEAMSGNDNDVYKQNNFSLSIDNDQSGMLKAFG 508
DQ 459 ISSDYLLKAAAGIDKNNKANNHVSIVEAWSNDNTPYLDHDDGNLMMNDRKFLSLMSLSLA 518
QY 509 YASAYRGNLSNLATAGLKNRSAMP--DSDPVPNVYFIRAHDSVQTRIAKIIREKLGKTN 567
DQ 519 KPLDKESGLNPLHNSLVREVDREVEITVPSYSFARAHDSVQDIIIRDIKAEI--NPRS 577
QY 568 DGLTNLTLDLNLKAFDIYNQDMNATDKVYYPNNLPMAYAMWLQNKDQTVTRVYVYGDMYTGN 627
DQ 578 FGYSFTQBEIEQAFKIYNEDLKTKKTHYVNVPLSYTLLLTNKGSIIPRVYVYGDFTDD 636
QY 628 GQYMAKTFPYNAIETLLKRIKRYAGGQAVSVKQDWSSGILTSVRYGKGNASASDAGNT 687
DQ 637 GQYMAKNTVNYDAIESLLKARMKYSVGQAMQNYQIENGELITSVRYGKALKQSDKGA 696
QY 688 ETRNSGMALLINRPNFRA--YRNLNLMGAHKSQAYRPLLSTKDGATYLNDSVDV--S 745
DQ 697 TTRTSQGVVMGNQPNFSLDGKVALNMGNANQYERALMVSTKDGVAITYATDADASKA 756
QY 746 ROYKYTDSQNLISFSAEISQSVANQVSGMIQWVPVGAADNQDVRTSPSTQATKQNTY 805
DQ 757 GLVKRTDENGILYFLNDDLKGVANPQVSGFLQVWVPVGAADDQDIRVAASDASTASTDGKSL 816
QY 806 HQSDALDSQVIVYEGFSNFAQAFQSPDOYTNNAVITAKNGDLFKSGITQFEMAPQYVSSDGG 865
DQ 817 HQDAAMDSRVMFEGFSNFQSFATKEBEYNNVJANNVDFVPSWGIITDFEMAPQYVSSDGG 876
QY 866 TFLDSVILNGYAFSDRYDLAMSKNNYKSGKQDLANAIKGLQSAIGIKVLSDLVNPQNLNLP 925
DQ 877 QFLDSVJQNGYATFDYDLGMSKANKYGTADQLVKAIKALHAKGLKVMADWVPQMYTFP 936
```

RESULT 9

```
US-10-797-821-37
; Sequence 37, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 37
; LENGTH: 1590
```



```
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-797-821-37

Query Match          41.9%  Score 3031; DB 5; Length 1590;
Best Local Similarity 41.6%  Pred. No. 8.4e-174;
Matches 672; Conservative 217; Mismatches 436; Indels 292; Gaps 36;

QY 1 MEKRLAKYKHKVQWVAIGVTTVT-LSFILAGQVVAADTNNNDGTSVQVKNKVPDPKF 59
DQ 1 MEKNVRFQHKVKVKKVWTVLSVASATMLASALGASVASADTDTASDSDNQA--VVTGD--- 55

QY 60 DAQAQNGOLAQAAMFKAANQADQATSQVSPATGCRVDNQVTPAANQPAANVADQV--- 115
DQ 56 -QTTNQADTQTSIAA-----TATSEQASASTDAATDQ--ASAAEQGTGTASTDTAAQT 106

QY 116 ---ANPA-----TDGALNRQSAADTSTDGKAVPQTSDDPGHLETVDGKTYV 160
DQ 107 TTNAEAKVPTENENQGFDEMLAEAKVATAESD--SIPSLAKNSNVKQVQDKYV 164

QY 161 DANGORLKNYSMVIDGKTYVFDGQTG-----EAQTDLPKTKQAANDNVPSQYQANQAYSN 216
DQ 165 DQDGNVKONFAVSGVKIYVFD-ETGAYKDTSKVDADKSSAVSQNA-TIFAANNEYAST 222

QY 217 EASSFETVDNYLTADSWYRPRKILKNGSQWQASBGLRPLMTWPDAAATKAAAYANFWA 276
DQ 223 SAKNFEAVDNYLTADSWYRPRKILKNGSQWQASBGLRPLMTWPDAAATKAAAYANFWA 282

QY 277 KEGILSGSY--RONSANLDAATONIOSAJEKKIASEGNTNMLRDKMSQFVKSONOWSIAS 334
DQ 283 KVGIDKTYAETSQADLTAAELVQARIEQKITSNNTKWLREASAFVKTQPOWNGES 342

QY 335 ENETVYPNODHMGQGGALLFSNSKD--TEHANSWRLNLRNPTFOTGK-QKYFTTN-----YA 388
DQ 343 EK----PYDHLQNGALLFDNQTDLTPTDQSNRYLLNRPTNQTGSLDSRFTYVNDPLG 398

QY 389 GYELLANDVNSNPVQAEQLNHLHYLMNWGDIVMGDKDANPDGVRVDAVDNVDNADLLQ 448
DQ 399 GYDFLLANDVNSNPVQAEQLNHLHYLMNWGDIVMGDKDANPDGVRVDAVDNVDNADLLQ 458

QY 449 IQRDYKAKYGTQDNEKNAIDHLSILEANGNDNDYVKDQNFSLSDNDQSRGMLKARG 508
DQ 459 ISSDYLKAAAGYDKNNKNANNHVSIVEAWSNDNTPYLHDDGDNLMNMNDRKFLSLMLWSLA 518

QY 509 YASVRGNLSNLATAGLNKRSANP--DSDPVPNVFIRAHDSSEVQTRIAKIIREKLGKTA 567
DQ 519 KPLDKBSGLNPLHLSLVREVDREVEVPSISFARAHDSSEVQDIIIRIIKAEI-NPNS 577

QY 568 DGLTNLTLDLNAKAFDIYNQDMNATDKVYYPNNLPMAYAMLQNDKTVTVVYVYGDMDYDN 627
DQ 578 FGY-SFTQBEIEQAFKIYNEDLKTKDKYTHYVPLSYTLTLLTNKGSIPRVYVYGDMDFTDD 636

QY 628 GOYMATKTPFYNAIETLLKGRIKYVAGGQAVSKQDWSSGILTSVRYGKANSASDAGNT 687
DQ 637 GOYMAKNTVYDAIESLLKARMKYVSGGQAMQYQIGNGEILTSVRYGKALKQSDKQDA 696

QY 688 ETRNSGALLINRPNFRA--YRNLTLNMGAAHKSQAYRPLLLSTKDGIAATYLNDSVD--S 745
DQ 697 TTRTSGVGVVMGNQPNFSLDGKVALNMGAAHANQYRALMSTKDGVIATYATDADASKA 756

QY 746 ROYKYTDSQGNLSFSAEISQSVANAQVSGMIQVWVPVGAADNQDVRTSPSTQATKDNTY 805
DQ 757 GLVKRTDENGILYFLNDDLGKVPANQVSGFLQVWVPVGAADQDQIRVAASDSTASTDGKSL 816

QY 806 HQSDALDSQVIVEGSNFQAFQSPDYNTAVIANKGDLFKSWGITQFEMAFQYVSSEDG 865
DQ 817 HQDAAMDSDRMFEGFSNFQSPATKEEYTNVVIANNVDVFXVSGWIGITDFEMAFQYVSSTDG 876

QY 866 TFLDSVILNGYAFSDRYDLWMSKNNKYQSKQDLANAIKGLQAGIKVLSDLVPNOLYNLP 925
DQ 877 QFLDSVILNGYAFSDRYDLWMSKNNKYQSKQDLANAIKGLQAGIKVLSDLVPNOLYNLP 936

QY 926 GKEVVTATRVNQYQAKSGATINKTPYVANTRSYG--DYOEQYGGKFLDLDLQKLYPRLFST 984
```

RESULT 10

```
US-09-740-274-2
; Sequence 2, Application US/09740274
; Patent No. US20020031826A1
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
```


Db	398	DNTGGYFLLANDVNSNPVQAEQUNLWHLFKNFGNIYANDPANFDSIRVDAVDNDV	457
Qy	444	ADLLQIQORDYYKARYGTDQNEKNAIDHLSILEAWSGNDNDYVKDQNNFSLSDNDORS	503
Db	458	ADLLQITAGDYLKAAGIHKNDKAANDHLSILEAWSNDNTPYLHDDGDNMINMDNKLRL	517
Qy	504	LKAFGYASAVRGNLSNLAATAGLKRS--ANPDSDPVNYVPIRAHDSQVTRIAKITREKL	562
Db	518	LFLSLAKPLNORSQGNPLITNSLNRDNDNAETAAPVSYSFIRAHDSQVQLIADIKAIE	577
Qy	563	GKTNADGLTNLTDLNLKAFDIYNQDMNATDKVYYPNNLPMAIYMWLQNKDQTVRVVYGD	622
Db	578	NPNVGY--SFTWEEIKAEIYNKOLLATEKKYTHYNTALSYALLTNKSSVPRVYGD	635
Qy	623	MYTDNGOYMATKTPFYNAIFTLKGRIKYVAGQAVSYKODWSGILTSTVRYGKGANSAS	682
Db	636	MFTDDGOYMAHKTINYEAIETLLKARIKYVSGQAMRNQOVGNSEIITSVRYGKGALKAT	695
Qy	683	DAGNTEPRNSGMALLINNRNFR--AYRNLTANMGAHKSOARPILLSTPKDGIATYLN	740
Db	696	DTGDRITRTSGVAIEGNNFSLRLKASDRVVVNMGAHKMQARPILLTTDNGIKAYHSD	755
Qy	741	SDVDSROYKYTDSQGNLSFSAELOSANAAQVSGMIQVWYPVGA--DNQD	789
Db	756	QEA--AGLVRVTNRGELIFTAADIKGYANQVSGYLGWVPVGAALIKMPALRLARPHQ	814
Qy	790	VRTSPSTQATKGNIIYHQSALDSOVIYEGFSNFOAPAQSPDOYTNATVIAKNGDLFKSWG	849
Db	815	MAS-----VHONAALDSRVNFEFSNFOAPATKKEEYTNVVIKAVNDKFAEWG	862
Qy	850	ITOPEMAPQVVSSEGGTFLDSVTLNGYAFSDRVDLAMSKNKYGSKQDLANAIKGLQSAG	909
Db	863	VTDFEMAPQVVSSTDSGFLDSVLQNGYAFDTRVDLIGISKENKYGTADDLVKAIKALHSG	922
Qy	910	IKVLSLIVPNQLYNLPKGVVTVATRVNQYGOAKSGATINTPTVANTRSYG-DYQSQYGG	968
Db	923	IKVMADWVPQMYAFPEKEVTVATRVVDKYTYPVAGSQIKNTLVVVDGSKSSGQQAQYGG	982
Qy	969	KFLDLOKLYPLFSTKQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKYVLSE--GNK	1026
Db	983	AFLEELQAKYPELFPARKQISTGVPMDDPSVKIKQWSAKYFNGTNILGRGAGYVLKQATNT	1042
Qy	1027	YLANLADGK--LFLPTVLNNTYGPQVPSANGFISKNGGIIHYLDKNQEVKNRFKEISGSWY	1084
Db	1043	YFNISDNKEINFLPKTLN-----QDSQVGSFYDGGKYVYVYSTSGYQAKNTFISEGDKWY	1097
Qy	1085	YFSDGKMAKTKIENDTVLFWPBGKOLKEGVW--YDGLKAYKYDDNGRGTWTKGFVEP	1142
Db	1098	YFDNNGYMWYGAOSINGVNTYFISNGIQLKDALIKNEDGYIAYTGN-DGRYEN-GYQF	1155
Qy	1143	RVDQDKWRYFNGDGTIAIGLVSLDNRTLYPDAYGYQVKQTV--TINGKSYTFDADQGD	1201
Db	1156	M---SGVWRHFN--NGEMSVGLTVIDGQVQYFDEMGYQAKGFVTADGKIRYFDKQSGNM	1211
Qy	1202	VQTDNANPACQOQAG--WKLGLDNQWGRKDGQLLTGEOITDGGKVPFDQNGVQVKGAT	1260
Db	1212	YR----NRPTIENBEGKLYLGE-----DCAAVTGSQTINGQHLHYFRANGVQVKGFEVT	1260
Qy	1261	DASGLVLFYDRDQGHQVKGKMYSTSDNNVYVNESGOVLGTLOITDQGTVYFDDKGIQAK	1320
Db	1261	DHGRISYIDGNSGDQIRNFRVNAQCQWYFDDNGYAVTGARTINGQLLYFRANGVOVK	1320
Qy	1321	GKAVWDENGLRYFPDADSGNMLDRW--KNVDGNWYFNRNGLA	1362
Db	1321	GEFVTRDYGRISYIDGNSGDQIRNFRVNAQCQWYFDDNGYA	1363

Db 1116 KNAFISLGNWYFDNNGWMTGAQASINGANYFSLNSGIQLRNAIYDNGNKVLSYYGNDG 1175
Qy 1132 RTWTNKGFEFRVDGQDKWRYFNNGDGTIAIGLVSLDNRTLIFYDAYGYQVKGTVTINGKS 1191
Db 1176 RRYEN-GYLYF--GQ-QMRYPO-NGIMAVGLTRVHGAVQYFDASGFGQAKGQFI----- 1223
Qy 1192 YTFDADQGLVQTDNANPAPQOAGWKLLGDNQWGYRKDGQLLTGEGTIDGOKVFPQDNG 1251
Db 1224 ----- 1223
Qy 1252 VQVKGGTATDASGLRFRFYDRDOGHQVGKGMYSTDDNWWVYVNESGOVLTGLQITIDGQTVY 1311
Db 1224 -----TTADGKLYRFDSDSGNQISNRFVNSKGEWFLFDHNGVAVTGTVTFNQRLY 1275
Qy 1312 FDDKGIQAKGKAVWDENGNLRYFDADSGNMLRDW-KNYDGNWYFNENGLA 1362
Db 1276 FKPNGVQAKGEFIRDANGYLRYVDPNSGNEVRNRFVNSKGEWFLFDHNGIA 1327

Search completed: February 11, 2006, 20:57:20
Job time : 133.176 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 20:42:47 ; Search time 7.76765 Seconds
(without alignments)
2306.008 Million cell updates/sec

Title: US-10-797-821-39
Perfect score: 7230
Sequence: 1 MEKNLYKLHKVKQWVAIG.....WKNVDGNWYFNRNGLATRW 1365

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 97014 seqs, 13122538 residues

Total number of hits satisfying chosen parameters: 97014

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA_New.*
1: /cgn2_6/prodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/prodata/1/pubpaa/US06_NEW_PUB.pep.*
3: /cgn2_6/prodata/1/pubpaa/US07_NEW_PUB.pep.*
4: /cgn2_6/prodata/1/pubpaa/PCT_NEW_PUB.pep.*
5: /cgn2_6/prodata/1/pubpaa/US09_NEW_PUB.pep.*
6: /cgn2_6/prodata/1/pubpaa/US10_NEW_PUB.pep.*
7: /cgn2_6/prodata/1/pubpaa/US11_NEW_PUB.pep.*
8: /cgn2_6/prodata/1/pubpaa/US60_NEW_PUB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	367.5	5.1	2710	7 US-11-051-453-41	Sequence 41, Appl
2	334.5	4.6	2367	7 US-11-051-453-42	Sequence 42, Appl
3	290	4.0	1155	6 US-10-793-626-1780	Sequence 1780, Ap
4	280	3.9	726	7 US-11-052-554A-217	Sequence 217, App
5	270.5	3.7	690	7 US-11-052-554A-232	Sequence 232, App
6	248.5	3.4	396	7 US-11-022-562-228	Sequence 228, App
7	245.5	3.4	1992	7 US-11-013-759-3	Sequence 3, Appli
8	245.5	3.4	1992	7 US-11-013-759-13	Sequence 13, Appl
9	245.5	3.4	2047	7 US-11-013-759-4	Sequence 4, Appli
10	245.5	3.4	2047	7 US-11-013-759-7	Sequence 7, Appli
11	244.5	3.4	701	7 US-11-052-554A-231	Sequence 231, App
12	242.5	3.4	2314	7 US-11-013-759-11	Sequence 11, Appl
13	240.5	3.3	619	7 US-11-052-554A-229	Sequence 229, App
14	238.5	3.3	1290	6 US-10-485-517-141	Sequence 141, App
15	231	3.2	2053	7 US-11-013-759-9	Sequence 9, Appli
16	229.5	3.2	744	6 US-10-873-528-184	Sequence 184, App
17	228.5	3.2	332	6 US-10-873-528-190	Sequence 190, App
18	226.5	3.1	5024	6 US-10-793-626-2964	Sequence 2964, Ap
19	218.5	3.0	340	7 US-10-873-528-188	Sequence 188, App
20	216.5	3.0	1588	7 US-11-052-554A-280	Sequence 280, App
21	216	3.0	2399	7 US-11-052-554A-92	Sequence 92, Appli
22	213	2.9	658	6 US-10-873-528-17	Sequence 17, Appl
23	213	2.9	677	6 US-10-873-528-155	Sequence 155, App
24	212.5	2.9	1263	6 US-10-485-517-127	Sequence 127, App
25	212	2.9	693	6 US-10-873-528-185	Sequence 185, App

ALIGNMENTS

RESULT 1

US-11-051-453-41
; Sequence 41, Application US/11051453
; Publication No. US20050287150A1
; GENERAL INFORMATION:
; APPLICANT: AMBROSINO, DONNA
; APPLICANT: BABCOCK, GREGORY J.
; APPLICANT: BROERING, THERESA
; APPLICANT: GRAZIANO, ROBERT
; APPLICANT: HERNANDEZ, HECTOR JAVIER
; APPLICANT: LOWY, ISRAEL
; APPLICANT: MANDELL, ROBERT
; APPLICANT: MOLRINE, DEBORAH
; APPLICANT: THOMAS, JR., WILLIAM D.
; APPLICANT: ZHANG, HUI-PEN
; TITLE OF INVENTION: ANTIBODIES AGAINST CLOSTRIDIUM DIFFICILE TOXINS AND
; FILE REFERENCE: MJ1-001
; CURRENT APPLICATION NUMBER: US/11/051,453
; CURRENT FILING DATE: 2005-02-04
; PRIOR APPLICATION NUMBER: 60/542,357
; PRIOR FILING DATE: 2004-02-06
; PRIOR APPLICATION NUMBER: 60/613,854
; PRIOR FILING DATE: 2004-09-28
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: Patentin Ver. 3.3
; SEQ ID NO 41
; LENGTH: 2710
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-11-051-453-41

Query Match 5.1%; Score 367.5; DB 7; Length 2710;
Best Local Similarity 19.4%; Pred. No. 1.2e-13;
Matches 265; Conservative 194; Mismatches 507; Indels 401; Gaps 65;
QY 240 LKNGOSQWASSSGDLRPILMTWVDAAT-KAAYANFAKEGLISGSYRQNSANLDAATQN 298
Db 1309 IRNKLSYFDGAGGYVLSLLSYPISTNLSKDDLWI-----FNIDNEVRE 1355
QY 299 IQSAIEKTIASRGNTNWLDRKMSQFVKSONQWSIASENETV-----YPNQDH----- 345
Db 1356 I--SIENGTFKKGKL--IKDVLKIDINKKLIIG--NOTIDFGSIDNKRVIYFUTCEL 1409
QY 346 -----MOGGALLFNSKDTTEHANDRLNRPFTQTKQKFTTNYAGYEL 392
Db 1410 DDKISLIIBINLVAKSYSLLLSGDK-----NYLISLSNTIE-----KINTLGLDS 1455

Sequence 191, App
Sequence 90, Appl
Sequence 228, App
Sequence 5, Appli
Sequence 8, Appli
Sequence 30, Appl
Sequence 372, App
Sequence 260, App
Sequence 91, Appl
Sequence 281, App
Sequence 1, Appli
Sequence 3154, Ap
Sequence 2, Appli
Sequence 211, App
Sequence 212, App
Sequence 45, Appl
Sequence 27, Appl
Sequence 28, Appl
Sequence 25, Appl
Sequence 95, Appl

Db 1625 ICENNNIQYFTKNTLETNYLYVGNQNMIVEPNYDLD-----DSGD 1669
QY 733 GIATYLDSDVDROQYKTDQGNLSPSASELOSANAQVSGMIQVWPVGAADNDQVRT 792
Db 1670 ISSTVIN---FQKILY-----GIDSCNVKV-----I 1694
QY 793 SPSTQATKQNIYHQSDALDSQVIYEGFSNFQAFQSPDOYTNVIAKNGDLFKSWGITQ 852
Db 1695 SP-----NIV--TDEINITPVYETNTNYPEVILVDANYINEKINVIN----- 1735
QY 853 FEMAPQVVSSEDG-----TFDLSVILN--GYAFSDRYDLAMS 887
Db 1736 -DLSIRVYDNGDNDFLMSTSEBENKVSQVIRFVNVFKDTLANKLSFNFSQKQVPS 1794
QY 888 KNNKYPQSDLANAIGLQAGIKVLSDLVPNOLYNLPKGEVVTATRVNOYQGAQSGAT- 946
Db 1795 -----EILSFTSYIEDGL-IGVDLGLVLYN-----EKPYINFGVMVGLIY 1838
QY 947 INKTPY-----VANTRSYGDOYQYGGKFLDDLQKLYPRLPSTKQISTGKPIDPSV 997
Db 1839 INDSLYYFKPPVNNLITGFVTGD-----DKYFNPINGGAASIGETI---I 1882
QY 998 KITWBAKYPNGSNILRGAKYVLSBGNKYNLADGKFLP--TVLNNTYQOVQVANGF 1055
Db 1883 DKQY---YFNQSGVLQTV-FSTEDGFKY-----FAPANTLDENLEGE-AIDFTGK 1929
QY 1056 ISKNGGHIYLDKNGQVNRKFKISGSWYVYFSD-GKMATGKTIGNDTYLFWPCKQK 1114
Db 1930 LIIDENIYYPEDNYRGA-VWKVGYTE--IDGKHVFAENGEMQIGVENTEDGKYFAH 2043
QY 1115 EGVYDQKAYYDDNGRTWTKNGFVEFRVDG-----QDKWRYF--- 1153
Db 1989 GFVSINDNK-HYFDDSG--VMKVGYTE--IDGKHVFAENGEMQIGVENTEDGKYFAH 2043
QY 1154 -----NGDGTAGLVDNRTLYFDAYGYVQGTQVTTINKSYTFDADQDLVQTON 1206
Db 2044 NEDLNGEETISYSILNFKNIIYFDDSFYAVVGWKOLEDGSKYFYFDED----- 2094
QY 1207 ANPAPQAGWKLGNQYRKGOLLTGEQITDGQVFFQDNGVQVKG----- 1256
Db 2095 ---TAEAYIGLSINDQYIFNDGIMQVGFVTTINDKVFYFSDGIESGVQVNDIDNYFY 2151
QY 1257 -----GTATDASGVLR-----FYDRDQGHQVQKG 1280
Db 2152 IDDNGIVQIGVPTSDGYKYFAPANTVNDNIYQAVEYSGLYRVGDEVYFGETYIETG 2211
QY 1281 W-----YSTDDNWWYVNESQVLTGLQITDQTVYFDDKGIQAKGKAWDENGNLRYFDAD 1337
Db 2212 WIYDMENESDKYFVPETTKACKGINLIDDIKYFDEKGMRTGLISFENN---NYIFYNE 2268
QY 1338 SGNMLRDRKNDGNWYFNRLGL 1361
Db 2269 NGE-IPFGYINIEDKMFYFGEDGV 2291

RESULT 3

US-10-793-626-1780
; Sequence 1780, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1780
; LENGTH: 1155
; TYPE: PR

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-1780

Query Match 4.0%; Score 290; DB 6; Length 1155;
Best Local Similarity 19.9%; Pred. NO. 1.4e-09;
Matches 287; Conservative 185; Mismatches 547; Indels 422; Gaps 66;

QY 1 MEKRLYKLYKHYKQWVAIGVITVTLISFLAGGVV--AADTNNDGTSVQVNMVSPDPK 58
Db 9 INKRVDFLGNKVNKYSIR-KFTVGTASILVGAATLFGAADNEAKAAEDNQLESASKEEQK 67
QY 59 FDAQNGQLAQAFKAAQADQATATSOVSPATD-GRVDNQVTPAANQAPAA--VANQD 114
Db 68 GSRDNESKLNQVDLNGSHSSE-KTNNVNNATEVKKVEAPTSDVSKPKANEAVTNES 136
QY 115 VANPATDAGALNRQSAADTSTDGKAVPQTSQDQGHLETVGKTYVYDANGQRLK---NYS 171
Db 127 TKPKTTEAPTVEESIAET-----PKTS-----TTQDST---EKNPSSLKDNLS 170
QY 172 MVIDGKTYFDCGTGEAQTDLPKTGOANQDNVDPDSQOANNOAYSNEASSFETVDNLTAD 231
Db 171 STTSKESKTDHSTKQAQWSTNKNLDTNDSPQSEKTSQANN-----DSTDN----- 219
QY 232 SWYRPRKILKNGQSQASSEGDLRPLMTWMPDAATKAAVAFWAKEGLISGSYRQNSAN 291
Db 220 -----QSAPSKQLDSKPEQKYTKFNDPTQDVEHT-----TTKLTPTPSIS 262
QY 292 LDAATQNIQSAIEKTIAS-----EGNTNLRDKMSQFVKSQ---NQWSIASENETVY 340
Db 263 TDSSVNDKQDYTRSAVASGLVDSNETEAITNAVNDLNLKAASREQINEAIIAEALKKDF 322
QY 341 PNQDHMQGG-----ALLFSNKDTEHANSDWRL-----LNRNPTFQTKQKYFTTNVAGYE 391
Db 323 SNPDY---GVDTPALNTSQSKNSPHKSASPRMLMSLAEP---NSGKN----- 366
QY 392 LLLAND-VDNSNPVQAEOLNHLHYLMNMGDTVMGDKDANFDGVRVDAVDNVNADLLQIQ 450
Db 367 ---VNDKVKITPTLSLKNSSN---HANNVWPTSNEQFN----- 400
QY 451 RDIYKAKYTDQNEKNAIDHLSILEAWSGNDNDYVKDQNNFSLSIDNDORSGLMAFGVA 510
Db 401 ---LKANYELD-----DSIKEGDTFTIKYQYIRPGGLELPAIK 436
QY 511 SAYRGNLSMLATAGLNKRSANPDSVPVNYVFIKRAHDSVQVTRIAKIIREKLGKTNADGL 570
Db 437 TQLRSKQSGIVANGVYDKTNTTNTTNYVDQYQNIITGSFDLIATPKRETAIKDNQNP 496
QY 571 TNLTL--DDLKAF--DIYNQDMNAT-----DKVYYPNNLPMAVAMWLQNKDVTVRVY 619
Db 497 MEVTIANEVVKKDFIVDGNKNDNTTAAVANVDV---NNKNEVVYLNQNNQNPYAK 553
QY 620 YGDMYTDNGQYMATKPPFYNAIET-----LLKGRIKYVAGGQAVSYKQDWSGSL 669
Db 554 YFST-VKNGKPIPGEVKVEVTDITNAMVDSFNPDLNSSNVKDVTSQFTPKVSAD---GTR 609
QY 670 TSVRYGKG-ANSASDAGNTETRNSGALLINRNPFRAYRNLTLNM-GAAHKSQAVRPLL 727
Db 610 VDNIFARSMAKGIIVTQAVRPTGTG-----NVYTEWLTTRDGTNTNDFV--- 657
QY 728 LSTKQGIATYLDSDV---DSQYKYTDSQGNLSFSASELSQVANAQVSGMQLVMPVGA 784
Db 658 -GKSTTTVTLNGSSTAQGDNPY-----SLGDYVWLDKKNKNGVQDDDEKGLAGVYVTLKD 712
QY 785 ADNQDVRTSPSTQATKQNIYHQSDALDSQVIYEGFSNFQAFQSP-DQYTNVIAKNGD 843
Db 713 SNNRELQ-----RVTTDQSGHYQFNLQNGTYTVFEAFIPDNTYTPSPANNSTDAIDSDGE 767
QY 844 LFKSWGITQFENAPQVSVSEDTGTFDLSVILNGYAFSDRY---DLAMSKNKYKSGKODLAN 900
Db 768 ---RDGTRKVVVAKGTINNADNMTVDT-----GFYLTPKYNVGVYVWEDTKDGIQDD--- 817

QY 901 AIKGLQSAGIKVLSDLVPLNQLYHLPKGEVVTATRVNOYGOAKSGATINKTPYVANTRSYG 960
Db 818 -----NEKGISNVKVLKN-----KNGDTIGTT-----TTDSNG 846
QY 961 DYQEQYGGKFLDDLQKLYPRLFTSKQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKYV 1020
Db 847 KYE-----FTG----- 852
QY 1021 LSEGNKYLNADGKLFPTVLNNTYQPOVSANGFISKNGGIHYLDKNGQEVKNRPFKEIS 1080
Db 853 LENGDTIBETPEGYTPTKQN-----SGSDSGKDSNGIKTIVTVK----- 893
QY 1081 GSWYFDSGKMATGKTIGNDYLFMPNGKQLKEGWDGKAYYYDDNGRTWTKNGFV 1140
Db 894 -----DADNKKTI-----DSGFYKPI-YNLGDYVWEDTNKDGIQDD-----SEKGIS 933
QY 1141 EPRVDGDKWRYFNGGTTAIGLVSLDNRPLYDAYG-YQVKGQTVTINGKSYTFD- 1195
Db 934 GVKVTLUKDK-----NGN-----AIG-----TTTTDASGHYQFKG-----LENGSYTVEFETP 975
QY 1196 -----ADGGLDVQTDNANPAPQG-----QAGW-----KLLGDNQW-GYRKDGQ 1232
Db 976 SGYTPTKANSQGDITVDNSGIIITTTGIINGADNLITDSGFYKTPKYSVGVDYWEDTNKDG 1035
QY 1233 LLTGEOTIDGQKVFQDNGVQVKGATDASGLRVLYDRDQGHV-----GKGWYST----- 1284
Db 1036 QDNEKRGISGVKTLKDEKGNIIITTTTDENGKYQFDNLDNGNYIIHFKEKPEGMTQTAN 1095
QY 1285 --SDDNWWYVNESGQVLTGLQTDGQTVY-----PDDKGIQAKGKAVWDENGNLRYFDAD 1337
Db 1096 SGNDDE-----KADAGEDVRVTITDHDDESI-----DNG-----YFDDD 1129
QY 1338 S 1338
Db 1130 S 1130

RESULT 4

US-11-052-554A-217
; Sequence 217, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; PRIOR FILING DATE: 2005-02-07
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 217
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Streptococcus mutans UA159
US-11-052-554A-217

Query Match 3.9%; Score 280; DB 7; Length 726;
Best Local Similarity 34.1%; Pred. No. 2.9e-09;
Matches 62; Conservative 30; Mismatches 66; Indels 24; Gaps 5;

QY 1193 TEDADQGLVQTDNANPAP-----OQAGWKLLGDNQWGYRKDGQLLTGEQITDQ 1243
Db 127 SFKQQAHAHVKMTTEAKTPSHSINTFVNDGNGWYILG-----ADGRNVTGSHITGGK 179
QY 1244 KVFQDNGVQVKGATDASGLRVLYDRDQGHVQVKGW-----YSTSDNWWYVNESGQVLT 1300
Db 180 TMYFAQDGKQVKGAFQAQSDSGNKHYYDRDSGEN-----WTRNFVNDQGNWYLLNNDGVPVT 235

QY 1301 GLQTDIGQTVYFDDKGIQAKGKAVWDENGNLRYFPADSGNMLDRWKNVGDGNWYFNRRNG 1360
Db 236 GSITVNGSLYFNSDGSQVKGNFV-EEDGSLRYDKNSGDLURKTSRTINGVYFDNDG 294
QY 1361 LA 1362
Db 295 NA 296
RESULT 5
US-11-052-554A-232
; Sequence 232, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 232
; LENGTH: 690
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae R6
US-11-052-554A-232

Query Match 3.7%; Score 270.5; DB 7; Length 690;
Best Local Similarity 20.7%; Pred. No. 9.9e-09;
Matches 133; Conservative 102; Mismatches 220; Indels 189; Gaps 29;

QY 743 VDSRQKYKTDSSQNLSP-SASELQSVANAQVSGMIQVVPVGAADNQDVRTSPSTQATKD 801
Db 188 IGTSAPFSQKLUKLTFFSSSKLELISHEAFANL-----SNLEKLTLPKSVKTLG 237
QY 802 GNIYHQSDALDSQVIEGFSNFQA-----FAQSPQYTNNAVIAKNGDLFPKSWGITQFEMA 856
Db 238 SNLFRLTSLKHVDVEEGNESFASVDGVLFSDKDTQLIYVPSQKNDES YK----- 287
QY 857 POYVSEBDGTFDLSVILNGYAFSDRYDLAMSKNNKYGSKQDLANAIGKLSQAGIKVLSDL 916
Db 288 -----TPKETKELASYSFN-----KNSYLKKELEL-----EGLEKIGTFAFADA 326
QY 917 VPNQLYNLPKGEVVTATRVNOYGOAKSGATINKTPYVANTRSYGDYQEQYGGKFL-----D 972
Db 327 IKLEBISLP-NSLETIERLAFYGNLE-----LKEILILPDNVKNPGRHVMNGLPKFLTLSGN 391
QY 973 DLQKLYPRLFTSKQISTGKPIDPSVKITNWSAKYFNGSNILGRGAK-YVLSEGNKYL----- 1028
Db 382 NINSL-PSFFLSGVLDSLKEIHIKNKSTEFVYK--KOTFAIPEYKVFYVTSHEIKDVLKS 438
QY 1029 NLADGKLFPTVLNN-----TYGQPOVSAN-----GPISKNGGIHYLDKNGQEVKNRPFKEI 1079
Db 439 NLSTGNDIIVEKVDNIKQETDVAKPKNSNQGVGWVXDKGLWYLYNESGSMATGWVKD- 497
QY 1080 SGWYFDSGKMATGKTIGNDYLFMPNGKQLKEGWDGKAYYYDDNGRTWTKNGF 1139
Db 498 KGLWYLYNESGSMATGWVK-----DKGLW-----YLYNESGSMAT--GW 534
QY 1140 VEFVRDQDQK--WRYFNGGDTTIAIGLVSLDNRNLTYFDAYGYQVKGQTVTINGKSYTFDAD 1197
Db 535 V-----DKGLWYLYNESGSMATGWVK-----D 557
QY 1198 QGDLVQTDNANPAPQAGWKLLGDNQWGY-RDKGQLLTGEQITDQKVFQDNGVQVKG 1256
Db 558 KG-----LWYLYNESGSMATG-----WVKDKGL----- 580
QY 1257 GTATDASGLRVLYDRDQGHVQVKGWYSTSDNWWYVNESGQVLTGLQTDIGQTVYFDDKG 1316

Db 581 -----WYLNSSGSMATGWK-DKGLWYLNSSGSMATGWKDKGLWYLNSSG 628
Qy 1317 IQAGKAVDNGNLRFDADGNMLRDRKNDGNNYFNNG 1360
Db 629 SMATG---WYDKGLWYLNSSGSMATG-WYVSGKYYTNSG 668

RESULT 6

US-11-022-562-228

; Sequence 228, Application US/11022562

; Publication No. US20050249742A1

; GENERAL INFORMATION:

; APPLICANT: Ruprecht, Ruth M.

; APPLICANT: Shisong, Jiang

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING

; FILE REFERENCE: DPN-043CN

; CURRENT APPLICATION NUMBER: US/11/022,562

; CURRENT FILING DATE: 2004-12-22

; PRIOR APPLICATION NUMBER: PCT/US03/20322

; PRIOR FILING DATE: 2003-06-27

; PRIOR APPLICATION NUMBER: 60/392718

; PRIOR FILING DATE: 2002-06-27

; NUMBER OF SEQ ID NOS: 340

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 228

; LENGTH: 396

; TYPE: PRT

; ORGANISM: Clostridium difficile

US-11-022-562-228

Query Match 3.4%; Score 248.5; DB 7; Length 396;
Best Local Similarity 24.1%; Pred. No. 9.1e-08;
Matches 102; Conservative 67; Mismatches 162; Indels 93; Gaps 21;

Qy 969 KFLDDLOKLPRLPSTKQISTGKIDPSVKITWNSAKYFNQSN-ILGRGAKYVLSGNYK 1027
Db 2 EYYPEIIVLPNPTFHKK---VNINLSSSPEYKWKSTE--GSDFILVR---YLESSNKK 51

Qy 1028 LNLADGKFLPTVLNNYGOPOVSAN-----GFISKNGGIHYLDKNGQEVKVR--- 1075
Db 52 IL---QKIRIKGLSNTKSFNKMISDFKIKLSLGYIMSN---FKSFENSELDRDHL 104

Qy 1076 -FKEISGSWYFSDCKMATGKTIGNDTVLFPNGKQLKEGVW--YDGKAYVYDDNGR 1132
Db 105 GFKIIDNKYYDYDEASKVLKGLININNSLFYEDPIESNLVTG-WQTINGKK-YIFDIN-- 160

Qy 1133 TWTNKGFBFRVDGQDKRYFNGDGTIAIGVSLDNRTLYF---DAYGYQVKGQTV--- 1185
Db 161 --TGAASTSVKII-NGKHFFVNNNGVMQLGVFKPDGFEYFAPANTQNNNIEQAIVYOS 217

Qy 1186 ---TINGKSYTFDADGDLVQTDNANPAPCGQNGKLLGDNQWYR-KDQLLTGSGTID 1241
Db 218 KFTLNGKKYFYDND------KAVTGMOTIDGKKYFNLTAEAATGWTID 264

Qy 1242 GOKVFPDNGVGVKGTATDASGLRFDYRQDQHQVG--KG-----WYSTSD----- 1286
Db 265 GKYYFNTN-TSIASCTYIINGKHFFYFNDGIMQGVFKGNGFYPAPANTDANNIEG 323

Qy 1287 -----DNWVYNSGQVLTGLQTDIGQTVTFDDKGIQAKGKAVDENGRLY 1333
Db 324 QAIRYQNRFLYLDHNIYYFGNNSKAVTGMQTINGNYYFMPDPTAMAAGGLFIDGVIYF 383

1334 FDAD 1337

384 FGVD 387

RESULT 7

US-11-013-759-3

; Sequence 3, Application US/11013759

; Publication No. US20050249747A1

; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; FILE REFERENCE: 1038-921MIS:jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1992
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-3

Query Match 3.4%; Score 245.5; DB 7; Length 1992;
Best Local Similarity 19.0%; Pred. No. 1.3e-06;
Matches 307; Conservative 182; Mismatches 574; Indels 549; Gaps 75;

Qy 5 LRYKLHKVKQWVAIGVTTVTLFSLAGGQVVAADTNNNDGTSVQVNMVPSDPKFDAAQ 64
Db 353 LKVKLAKTLNLTENVTTILN-----ATTTKVGSSTTAAELLSLSLTFTQPN 401

Qy 65 NGQLAQM-----PKAANQADQAT-----SQVSPATDGRVDNQVTP----- 101
Db 402 TGSQSTKTVYGVNGVKTNNAAETTAAGITRITRKIGFARDGVDEKQAPYLDKKQLK 461

Qy 102 -----AANQPAANVANQDVANPATDAGALNQSAADTSDGKAVPQTSQDPG 148
Db 462 VGSVAITDINGIDAGNKKISLAKGSSANDAVTIEQL--KAAPKTLNAGAGISVTPE-- 517

Qy 149 HLETVDGKTYVYDANGQRLKNTSMVIDGKTYTFDGTGEAQTDLPKTGAQNDQVPSYQ 208
Db 518 --ISVDAKSGNVT-----PTYNIGV-----KTTELNSDGTSDKFS 551

Qy 209 ANNQAYSNEASFETVDNYL-----TADSWYRPRKILKNGQWQASSEGDLRILMTWTP 263
Db 552 VRSGTNNSLVTAELASVLYNEVNRDADSAL-----QSFTVKEEDDDANAITVAK 602

Qy 264 DAATKAAYANF-----WAKEGLISGYRQNSA-----NLDAATQNTQSAI 303
Db 603 DTTKAGAVSILKLGKNGLTVAITKDGTVTFGLSQDSGLTIGKSTLNNDGLT--VKDTN 660

Qy 304 EK-KIASEG-----NTNW-----LRDKMSQFVKSQWQSIASENETVYPNQDH 345
Db 661 EQIQVANGIKFTNVNGSNPGTGIANTARITDKIG-FAGSDG-----AVDTNKPYLDDQK 715

Qy 346 MGGGALLFSNS-----KOTEHA-----N 363
Db 716 LQVGNVKITNTGINAGKRAITGLSPTLPSIADQSSRNIELGNTIQDKKSNAASINDILN 775

Qy 364 SDWRLNLRN-----PTFQTKQKQFTTYNAGYELLAND 397
Db 776 TGFNLKNNNPIDFVSTYDIVDFANGNATTATVTHDTANKTSKVYVDVNVDDTTIHLTGT 835

Qy 398 VDSNPVVAEQELNHLHLMWNG---DIVMGDKANFDGVRVDADVNVNADLLQIQDYV 454
Db 836 DDKKGLGVKTKLNTSANGNTATNPNVNSSDEDAVNA--KDIAENLT-----LAKEIH 889

Qy 455 KAKYGTQD-----NEKNVAIDHLSILEAWSGNDNDYV-----KQNNFSLSDNQ 499
Db 890 TTKGTADTALQTFVTKVDENNADDANAITVQGNANNQVNTLTILKGEN--GLNKTDK 947

Qy 500 RSGMLKAFGASAYRGNLSNLATAGIKNRSANPDSDPVNPYVFIHADSEVQ-----TRI 554
Db 948 NGTVTFGTNTTSLGKAGKSTLNDGGLSTK--NPTG-----SEQIQVADGVK 993

Qy 555 AKIIREKLGKTNADGLTNLTLDL-----NKAFD-----IYNQDMNATDKVYPPNLPWA 604

```
Db 994 AKVNNNGVAGIDGTTTRIDRIGTGNGSLDKSKPHLSKDGINAGK-----1043
Qy 605 YAMLQNKQTVTRVYGYDMYTDN-----GQYMATKTPFYNAIETLLK-----646
Db 1044 -----KITNIQSGETAQNSHDAVTGKIVDLKTELENKISSTAKTAQNSLHEFSVA 1094
Qy 647 ----GRIKYVAGQAVSKQDWSGILTSVRYGKANSASDAGNTEFRNSGMALLINRPN 703
Db 1095 DEQGN-NFTVSNPYSSYDTSKTSDDVIT-----FAGENGITTKVKNKGW-----RVG 1139
Qy 704 FRAYRNLT---LNMGAHRSQAVRPLLLSTKDGIATVLDNSDVSQYKTVDSQGNLSFS 760
Db 1140 IDOTKGLTTPKLTVCN-----NNGKGIIV-----JDSQ-----NQQVITGL 1175
Qy 761 ASELOSVANAAQSGMIQWVPVGAADNQDVRTSPSTQATKDGNIYHQSD-----ALDSQVI 816
Db 1176 SNTLANVTN-----DKGSVRT-----TEQGNIIKDEKTRAASIVDVL 1213
Qy 817 YEGFSNFOAPAQSPDQYT-----NAVIAK--NGDLFKSWGI-----TQFEMA 856
Db 1214 SAGF-NLQNGEAVDFVSTYDYNFADGNATTAKVTYDTSKTSKVYVDNVDDTTIEYK 1272
Qy 857 PQVVS-----SEDGTFILSDVILNGYAFSDRY----DLAMSKNKYKSGKDLANAIGLQ 906
Db 1273 DKXGLVKVTTLTSTGTGANKFALSNQATGDALVKASDIVAHLNLTSGDIQTAKGASQANN 1332
Qy 907 SAGIKVLSDLVNPQLNLPKEVYVTRVNVQYQAKSGATINKTPYVANTRSYGDIQEOY 966
Db 1333 SAG-----YVDADGNKVIYDSTDNKYQAKNDGTVDKTEVAKDKLVAQAQTPD 1381
Qy 967 GGFELDDLQKLYPRLFSTKQISTGKIDPSVKITNWSAKYFNKSGNLTGRGAKVYLSEGNK 1026
Db 1382 G-----TLAQM-----NVKSVINKEQVNDANK-----KQGINEDNAFVKGLEKAASD-NK 1425
Qy 1027 YLMADGKFLPTV--LNNTYGOQVSA--NGFISKNGGHIYLDKNGQEVKGRFKE----1078
Db 1426 TKWAA-----VTGDLNVAQTPLETFAGDTGTTAKKGLTETIKGGQDTNKLNTNNIG 1479
Qy 1079 -ISGSWYIFDSDG---KMATGKTKIGNDTYLFMPNGKQLKEGVYWDGKKAYYYDDNGRTW 1134
Db 1480 VWAG-----TDGFTVKLAKDLTNLSVN-----AGGTKIDDKGVSFVDSGQAK 1523
Qy 1135 TNKGFEFRVDGQKWRFYNGDGTGIALGLVSLDNRTLYFDAYGYQVKGQVTVINGKSYTF 1194
Db 1524 AN-----TFVLSANGLDLGKVKVINSVNGK-TK 1549
Qy 1195 DADQGLDVOTDNA-NPAPOQAGKLLGDNMWGYR-----KDGQ 1232
Db 1550 DTDAAVQQUNEVNRLGLGNAG-----NDNADGNQVNIADIKKDPNSGSSNRTVIAKT 1605
Qy 1233 LITGEQITDGQKVFQDNGVQV---RGGTAT--DASGLVRFYDRDOGHQVKGWYSTSDDN 1288
Db 1606 VLGGKGNNDTEKL--ATGGIQQGVQVQDGNANGDLSNV-----WVKTKDG 1648
Qy 1289 WTVVNESGQVLTGLQITDGTQVTFDDKGIQAKGKAVWDENGLRYFDADSGN 1340
Db 1649 -----SKRALLATYNAAGQTNLYTNPAEADRI--NEQG-IRFFHVNDGN 1691
```

RESULT 8

```
US-11-013-759-13
; Sequence 13, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN OF MORAXELLA
; FILE REFERENCE: 1038-921MIS:jb
; CURRENT APPLICATION NUMBER: US/11/013,759

; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 1992
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
; US-11-013-759-13

Query Match 3.4%; Score 245.5; DB 7; Length 1992;
Best Local Similarity 19.0%; Pred. No 1.3e-06;
Matches 307; Conservative 182; Mismatches 574; Indels 549; Gaps 75;

Qy 5 LRYKLHKVKQWVAIGVTTVTLSFLAGGQVAAADTNNDNGTSVQVNMKVPSPKPFDAQAQ 64
Db 353 LKVKLAKTLNLTVEVNTTLN-----ATTVKVSGSSSTTAELLSDSLTFQPN 401
Qy 65 NQLAQAM-----FKAANOADQAT-----SQVSPATDGRVDNQVTP-----101
Db 402 TGSQSTSKTVYGVNGVKFTNNAETTAAGTTRITRDKIGFARDGDVDEKOAPYLDKCKQLK 461
Qy 102 -----AANQPAANVANQDVANPATDAGALNRQSAADTSTDCGKAVPQTSDDPG 148
Db 462 VCSVAITDNGIDAGKKISNLAKGSSANDAVTIEQL--KAAPTLNAGAGISVTPTE--517
Qy 149 HLETVDGKTYVVDANGQRLKKNYSMVIDGKTYVFDGQGEAQTDLPKTGOANQDNVDSYQ 208
Db 518 --ISVDAKSGNVTA-----PTYNIGV-----KTTLENSDGTSDKFS 551
Qy 209 ANQAYSNEASSFETVDNYL-----TADSWYRPRKILKNGSQWQASSEDLPILMTWTP 263
Db 552 VKSGTGNLSLVTAETHLASYLINEVNRDADSAL-----QSFTVKEEDDDANAITVAK 602
Qy 264 DAATKAAVANF-----WAKEGLISGSYRQNSA-----NLDAATQNTQSAI 303
Db 603 DTTKAGAVSILKLGKNGLTVAATKDGIVTFGLSQDSGLTIGKSLTNNDGLT--VKDTN 660
Qy 304 EK-KIASEG-----NTNW-----LRDKMSQFVKSQNMQWSIASENETVYPNQDH 345
Db 661 EQIQVANGIKFTNVNGSNPGTGIANATARTDKIG-FAGSDG---AVDTNKPYLDDQK 715
Qy 346 MOGGALLFSNS-----KDEHA-----N 363
Db 716 LQVGNVKITNTGINAGGKAITGLSPTLPSIADQSSRNIELGNTIQDKKSNAAASINDILN 775
Qy 364 SDWRLNLRN-----PTFOTGKQKPYFTTVYAGVELLLAND 397
Db 776 TGFNLKNNNPIDFVSTYDIVDFANGNATTATVTHDTANKTSKVYDVNVDDTTHLTGT 835
Qy 398 VDSNPVVOAEQLNHLHYLMNWG---DIVMGDKDANFDGVRVDADVNNVADLIQIQDY 454
Db 836 DDNKKLGVKTKLNKTSANGNTATNFVNSSDEDALVNA--KDIAENLAT---LAKEIH 889
Qy 455 KAKYGTQD-----NEKNAIDHLSILEAWSGNDNDYV-----KQNNFSLSDNDQ 499
Db 890 TTKGTADTALQPTTVKVKVDENNNDADANAITYGQKNNANQVNTLTLLKGEN--GLNIKTDK 947
Qy 500 RSGMLKAFGYASAYRGNLSNLATAGLKNRSANPDSDPVPVYVFIRAHDSQV-----TRI 554
Db 948 NGTVTFGINTTSLGKAGKSLTNDGGLSIK--NPTG-----SEQIQVGADGVKF 993
Qy 555 AKIIREKLGKTNADGLTNLTDDL-----NKAFD-----IYNQDMNATDKVYPPNLPMA 604
Db 994 AKVNNNGVVGAGIDGTTRITRDEIGFTGNGSLDKSKPHLSKDGINAGK-----1043
Qy 605 YAMLQNKQTVTRVYGYDMYTDN-----GQYMATKTPFYNAIETLLK-----646
Db 1044 -----KITNIQSGETAQNSHDAVTGKIVDLKTELENKISSTAKTAQNSLHEFSVA 1094
Qy 647 ---GRIKYVAGQAVSKQDWSGILTSVRYGKANSASDAGNTEFRNSGMALLINRPN 703
```

Db 1095 DEQGN-NFTVSNPYSSYDTSKTSVDIT-----FAGENGITTKVKNQGV-----RVG 1139
QY 704 FRAYRNLT---LNMGAHKSQAYRPLLLSTKDGIAITYLNDSDVDSRQKYKTDQGNLSFS 760
Db 1140 IDQTKGLTTPKLTGVN-----NNGKGV-----IDSQ-----NQGNITGL 1175
QY 761 ASLQSVANAQVSGMIQVWVPVGAADNQDVRTSPSTQATKGNIIHQSD-----ALDSQVI 816
Db 1176 SNTLANVTN-----DKGSVRT-----TEQGNIRKDEKTRAASIVDVL 1213
QY 817 YEGFSNFOAQPSPDOVT-----NAVIK--NGDLFKSWG-------TQFEMA 856
Db 1214 SAGF-NLQNGEAVDFVSTYDVFADGNATTAKVYDDTSKTSKVYVVDVDDTIEVK 1272
QY 857 PQYVS-----SEDTGLSDVILNGYAFSDRY-----DLAMSKNKYKSGKDLANAIKGL 906
Db 1273 DKKLGVKTTLTSTGTGANKFALSNOATGDALVKASDIVAHLNTLSGDIQTAKGASQANN 1332
QY 907 SAGIKVLSLDVFNQNLNLPQKEVVTATRVNQQYQAKSGATINTPTVANTRSYGDYQBY 966
Db 1333 SAG-----YVDADGNKVIYDSTDNKYQAKNDGTVDKTEVAKDKLVAQAQTPD 1381
QY 967 GKGFLDLOKLYPRLFSTQISTGKPIDPSVKITNWSAKYFNGSNILGRCAKYLSEGNK 1026
Db 1382 G-----TLQAM-----NVKSVINKEQVNDANK-----KQINEDNAPFVKGLEKASD-NK 1425
QY 1027 YLNADGKFLPLTV--LNNYTGQPVSA--NGFISKNGGIHYLDKNGQEVKXRFKE----- 1078
Db 1426 TKNA-----VTVGLNVAQTPFLTFAGDTGTTAKKLGTLTKGQDITNKLTDNNIG 1479
QY 1079 -ISGSWYFSDG---KMATGKTIGNDTYLFMPNGKQKLGWYDQKAYYYDDNGRTW 1134
Db 1480 VVAG-----TDGFTVKLAKDLTNLSVN-----AGGTKIDDKGVSFVDSGQAK 1523
QY 1135 TNKGFEVRVGDQKWRYFNGDGTIAIGLVSLDNRPLYFDAYGVQYKQGVTLINGSYTF 1194
Db 1524 AN-----TPVLSANGLDGGKVISNVKG-TK 1549
QY 1195 DADQGLVTDNA-NPAPOQAGKWLGDQNGYR-----KDGQ 1232
Db 1550 DTDAAVQQLNEVRNLLGLNAG---NNDAGNQVNIADIKDPNSGSSNRTVIKAT 1605
QY 1233 LITGEGTIDQKVFPODNGVQV---KGGTAT-DASGVLFYRDRDQHQVKGWYSTDDN 1288
Db 1606 VLGGKGNNDTEKL--ATGIGIQGVDRKQGNANGDLSNV-----WVKQKDG 1648
QY 1289 WYVNESGVVLGTLQITDQTYFDDKGIQAKGKAWDENGNLRYFDASGN 1340
Db 1649 -----SKALLATYNAAGTNYLTNNPBAIDRI--NEQG-IRFPFVNDGN 1691

RESULT 9

US-11-013-759-4
; Sequence 4, Application US/11013759
; Publication NO. UB20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN OF MORAXELLA
; FILE REFERENCE: 1038-921MIS:jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; PRIOR FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2047
; TYPE: PR1
; ORGANISM: Moraxella catarrhalis

US-11-013-759-4

Query Match 3.4%; Score 245.5; DB 7; Length 2047;
Best Local Similarity 19.0%; Pred. No. 1.4e-06;
Matches 307; Conservative 182; Mismatches 574; Indels 549; Gaps 75;
QY 5 LRYLKHVKQWVAIGVTTVTLSFLAGGVAAADTNNNDGTSVQVNNQVWSPDKFDAQA 64
Db 408 LKVKLAKLTNNLTENVNTTLN-----ATTTVKVGSSSTTAELLSLSLFTQPN 456
QY 65 NQOLAQAM-----FKALNQADOTAT-----SOVSPATDGRVDNQVTP----- 101
Db 457 TGSQSTSKTVYGVNGVKFTNNAETTAAGTTRITRDKIGFARGDGVDEKQAPYLDKKQLK 516
QY 102 -----AANQAPAAVANQDVANPATDAGALNRQSAADTSTDGKAVPQTSQDP 148
Db 517 VGSVAITIDNGIDAGNKKISNLAKGSSANDAVTIEQL--KAAKPTLNAGAGISVTPE-- 572
QY 149 HLETVDGKTYVYDANGQRLKNYSMVJDGKTYYPDGTGEAQTDLPTKQANQDNVPSYQ 208
Db 573 --ISVDAKSGNVA-----PTYNIGV-----KTELNSDGTSDKFS 606
QY 209 ANNQAYSNEASSPFTVDNYL-----TADSWYRPRKILKNGQSQWASSEGDLRILMTWP 263
Db 607 VKSGTNNSLVTAHLASYLNEVNRADSAL-----QSFTVKEEDDDANAITVAK 657
QY 264 DAATKAAYANF-----WAKEGLISGYSRONSA-----NLDAATONIQSAI 303
Db 658 DITKAGAVSIILKKGKGLTVATKQGVTVTGLSQDSGLTIGKSTLNDGLT--VKDTN 715
QY 304 EK-KIASEG-----NTNW-----LRDKMSQFVKSQNQSIASENETVYPNQDH 345
Db 716 EQIQVANGIKFTNVNGSNPGTGIANTARITRDKIG-FAGSDG-----AVDTNKPILDQDK 770
QY 346 MOGGALLFNS-----KDEHA-----N 363
Db 771 LQVGNVKITNTGINAGGKAITGLSPLPSIADQSSRIELGNTIQDKKSNAAINDILN 830
QY 364 SDWRLLRN-----PTFQTGKQKPYFTTNVAGVYELLAND 397
Db 831 TGFNLKNNNPIDFVSTYDIVDFANGNATTATVTHDTANKTSKVYVDVDDTTIHLGT 890
QY 398 VDNSNPVVOAEQLNHLHYLMNWG---DIVMGDKDANFDGVRVDADVNNADLLQIQDY 454
Db 891 DDNKKLGKVTTKLNTKSANGNTATNFNVNSDEDALVNA--KDIAENLAT---LAKEIH 944
QY 455 KAKYGTQ-----NEKNAIDHLSILEAWSGNNDYV-----KOONNFSLSIDNDQ 499
Db 945 TTKGTADTALQTFYVKVVDENNADANAITYGQKNANNQVNTLTILKGEN--GLNIKTDK 1002
QY 500 RSGMLKAFGYASAYRGNLSNLATAGLKNRSANPDSPPVNYVYFIRAHDSEVQ-----TRI 554
Db 1003 NGTVTFGINTTSGLKAGKSTLNDGGLSIK--NPTG-----SEQIQVAGDVKF 1048
QY 555 AKIIREKLGKTNADGLTNLTLDL-----NKAFD-----IYNQDMNATDKVYPPNNLPMA 604
Db 1049 AKVNNVGVGAGIDGTTRITRDEIGFTGNGSLDKSKPHLSKDGINAGK----- 1098
QY 605 YAWMLQNKDVTVRVYGYDMYTN-----GQYMATKTPFYNAIETLLK----- 646
Db 1099 -----KITNIQSGEIAQNSHDAVTGGKIYDLKTELENKISSAKTAQNSLHEFSA 1149
QY 647 ---GRIKVVAGQAVYKQDSSGILTSVRYKGANASADAGNCTETRSQGNLSPFS 760
Db 1150 DEQGN-NFTVSNPYSSYDTSKTSVDIT-----FAGENGITTKVKNQGV-----RVG 1194
QY 704 FRAYRNLT---LNMGAHKSQAYRPLLLSTKDGIAITYLNDSDVDSRQKYKTDQGNLSFS 760
Db 1195 IDQTKGLTTPKLTGVN-----NNGKGV-----IDSQ-----NQGNITGL 1230
QY 761 ASELOSVANQVSGMIQVWVPVGAADNQDVRTSPSTQATKGNIIHQSD-----ALDSQVI 816
Db 1231 SNTLANVTN-----DKGSVRT-----TEQGNIRKDEKTRAASIVDVL 1268

1
QY 817 YGFSNFQAFAPSPDOYT-----NAVIAK--NGDLFKSWG1-----TQFEMA 856
DB 1269 SAGF-NLQNGEAVDFVSTYDTVNFADGNATTAKVTDTSKTSKYVYDVNVDTTIEVK 1327
QY 857 PQYVS-----SEDGTFLDVILNGYAFSDRY-----DLAWSKNNKYGSKDLANA1KG1Q 906
DB 1328 DKKLGVKTTTLTSTGTGANKFALSNOATGDALVKASDIVAHLNTLSGDIQTAKGASQANN 1387
QY 907 SAGIKVLSDLPVNLNPKGEVVTATRVNOYQQAQSGATINKTPYVANTRSYGDYQEQY 966
DB 1388 SAG-----YVDAGNKVIYDSTDNKYQAQNDGTVDKTEVAKDLVAQAQTPD 1436
QY 967 GGFLLDLOKLPRLPSTKQISTGKPIDPSVKITWSAKYFNGSNILGRGAKYVISEGNK 1026
DB 1437 G-----TLAQM-----NVKSVINKEQVNDANK-----KQGINEDNAFVKGLEKAAASD-NK 1480
QY 1027 YLNLADGKFLPTV--LANNYGOPOVSA--NGFISKNGGIHYLDKNGQEVKNRFE---- 1078
DB 1481 TRNAA-----VTVGDLNAVAQTPFLTFAGDTGTTAKKLGGETLTIKGGQDTRNKLTDNNTG 1534
QY 1079 -ISGSWYFDSQG--KMATGKTIGINDTYLFPMPNGKQLKEGVYDGGKAYYYDNGRTW 1134
DB 1535 VVAG-----TDGFTVKLAKDLTNLSVN-----AGGKIDDKGVSFVDSGQAK 1578
QY 1135 TNKGFEVRVDGQKWRYFNGDGTIAIGLVSLDNRTLYFDAYGYQVKGQTVTINGKSYTF 1194
DB 1579 AN-----TPVLSANGLDLGGKVISNVGKG-TK 1604
QY 1195 DADQGLVQTDNA-NPAPOGAQWKLGLGNQGYR-----KDGO 1232
DB 1605 DTDAAVNVQVLEVRNLLGLGNAG-----NDNADGNQVNIADIKKDPNSGSSNRVTIKAGT 1660
QY 1233 LITGEOTIDGQVFFQDNGVQV---KGGTAT-DASGLVRFYDRDQGHQVKGWYSTSDN 1288
DB 1661 VLGGKNNTEKL--ATGSIQVGVCDGNGANGDLSNV-----WVKTKDG 1703
QY 1289 WYVNESGQVLTLGLOTIDGQTVYFDDKGIQAKGAVMDENGNIIRYPDASGN 1340
DB 1704 -----SKKALLATYNAAGQTNLTNNPAEADRI--NEQG-IRFFHVNDGN 1746
RESULT 10
US-11-013-759-7
; Sequence 7, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN OF MORAXELLA
; FILE REFERENCE: 1038-921MIS:jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; PRIOR FILING DATE: 2004-12-16
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 2047
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-7

Query Match 3.4%; Score 245.5; DB 7; Length 2047;
Best Local Similarity 19.0%; Pred. No. 1.4e-06;
Matches 307; Conservative 182; Mismatches 574; Indels 549; Gaps 75;
QY 5 LRYKLHKVKQWVAIGVTTVTLSFLAGGVVAADTNNNDGTSVQVKNVPSDPKPAQAAQ 64
DB 408 LKVKLAKTLNLTENVNTTLN-----ATTTVKVGSSSTTAEILLDSLTFQPN 456

QY 65 NQLQAQAM-----FKAANQADQAT-----SQVSPATDGRVDNQVTP----- 101
DB 457 TGSQSTSKTVGVNGVKFTNNAETTAA1GTTTRITRDKGFARGDVDVDRQAQVYLDKKQJUK 516
QY 102 -----AANQPAANVANQDVANPATDAGALNRQSAADTSTDCGKAVPQTSDDPG 148
DB 517 VGSVAITTDNGIDAGNKKISNLAKGSSANDAVTIEQL--KAAPPTLNAGAGISVTPTE-- 572
QY 149 HLETVDGKTYVYDANGQRLKNYSMWIDGKTYTFDQGTQEAQTDLPKTPGAQNDPNVDSYQ 208
DB 573 --ISVDAKSGNVA-----PTYNIGV-----KTTLSNDSGTSKXFS 606
QY 209 ANNOAYSNEASFEFVDNVL-----TADSWTRPKILKNGQSQWQASSEGDLRPLMTWWP 263
DB 607 VKSGTNNSLVTAELASLYLNEVNRRTADSAL-----QSFTVKEEDDDANAITVAK 657
QY 264 DAATKAAVANF-----WAKEGLISGSYRQNSA-----NLDAATQNIQSAI 303
DB 658 DTTKNAGAVSILKLGKNGLTVATKDGTVTFGLSQDSGLTIGKSTLNNDGLT--VKDTN 715
QY 304 EK-KIASBG-----NTNW-----LRDKMSQFVKSQNWSIASENETVYPNQH 345
DB 716 EQIQVANGIKFTNVNGSNPGTGIANTARITRDKIG-FAGSDG-----AVDTNKPYLDDQK 770
QY 346 MOGGALLFSNS-----LDRKMSQFVKSQNWSIASENETVYPNQH 363
DB 771 LOVGNVKITNTGINAGKAITGLSPTLPSIADQSSRNIELGNTIQDKOKSNAASINDILN 830
QY 364 SDMRLNLRN-----PTFQTKQKQYFTTNYAGVELLLAND 397
DB 831 TGFNLKNNNPIDFVSTYDVIDVPFANGNATTATVTHDTANKTSKVYDVNVDTTIHLTGT 890
QY 398 VDNSNPVQAEOLNHLHYLMNWG----DIVMGDKANFDGVRVDAVDNVDNADLLQIQDY 454
DB 891 DDNKKLGKVTTLKNTSANGNTATNFVNSSEDAVNA--KDIAENLT-----LAKETH 944
QY 455 KAKYGTQ-----NEKNAIDHLSILEAWSGNDNDYV-----KQNNFSLSDNDQ 499
DB 945 PTKGTADTALQTFVTKVDENNADDANAITVQKNANNQVNTLTGKGN--GLNKTDK 1002
QY 500 RSGMLKAFGYSAYRGNLSNLATAGLKNRSANPDSDPVNYPVFIHRAHDEVO-----TRI 554
DB 1003 NGTVTFGINTTSLGKAGKSTLNDGGLSIK--NPTG-----SQIQVGADGVKF 1048
QY 555 AKIIREKLGKTNADGLTNLTDLL-----NKAFD-----IYNQDMNATDKVYPPNLPWA 604
DB 1049 AKVNNNGVVGAGIDGTTRITRDEIGFTGTNGSLDKSKPHLSKDGINAGGK----- 1098
QY 605 YAWMLQNKDVTVRVYGYDMYTN-----GQYMATKTPFYNAIETLLK----- 646
DB 1099 -----KITNIOSGEIAQNSHDAVTGGKTYDLKTELENKISSAKTAQNSLHEFSVA 1149
QY 647 ---GRIKVAGQAVSYKQDWSSGILTSVRYKGANASADAGNTEPRNSGSMALLINRPN 703
DB 1150 DEQGN-NFTVSNPYSDTSKTSQDVT-----FAGENGITTKVNGGV-----RVG 1194
QY 704 FRAYRNLT---LNMGAHKSQAYRPLLLSTKDGIAITYLNDSDVDSROYKYTPSQGNLSPS 760
DB 1195 IDQTKGLITPKLTJVG-----NNGKGIV-----IDSQ-----NQNTITGL 1230
QY 761 ASELOSVAQAQVSGMIQVWPVGAADNQDVRTSPSTQATKDNIIHQSD-----ALDSQVI 816
DB 1231 SNTLANVTN-----DKGSVRT-----TEQGNIIKDEKTRAASIVDVL 1268
QY 817 YGFSNFQAFAPSPDOYT-----NAVIAK--NGDLFKSWG1-----TQFEMA 856
DB 1269 SAGF-NLQNGEAVDFVSTYDTVNFADGNATTAKVTDTSKTSKYVYDVNVDTTIEVK 1327
QY 857 PQYVS-----SEDGTFLDVILNGYAFSDRY-----DLAWSKNNKYGSKDLANA1KG1Q 906
DB 1328 DKKLGVKTTTLTSTGTGANKFALSNOATGDALVKASDIVAHLNTLSGDIQTAKGASQANN 1387

Db 434 TA-----EIQSGGLTFT-----PTTNASTDKTVYGTGDKFTDMSNTALEDTTRITKDKIGF 485
QY 385 TNYAGYELLANDVNSNPVQAEQUNHLHYLMNWG-----420
Db 486 SNKAG-----TVDENKPYLDKDKLVGNSTLNGLTVNNTIGGSNKQIQVGADGKIF 538
QY 421 -----DIVMGDKDANFD-----GVRV-----436
Db 539 ADVNVNVSNAKFGTTRITEEIEGFADAGKVDKKSPLYDKKQLOVGGVKVITKDSGINAG 598
QY 437 -----DAVNUNA-----DILQIORD-----454
Db 599 DOKISNVKQATDDTDAVTVYKQLKQVQDADGALQSPSIRDEKGOEFTINLNSNGWTPNT 658
QY 455 -----KAKYGTQ-----NEKNAIDHLSLEAWSGND 481
Db 659 FETITPAGENGISINDIAKGVKVGIDPINGLTTPKLTGSDKDGKTQLVIEQVASGND 718
QY 482 NDYV-----KOQNFSLSIDNDQSRGML-----KAFG 508
Db 719 TKNIIRGLSPTLPSITNAGGVRTTEQGNITTSDEDKSAAISGDIILNTGFNLKNSNSVG 778
QY 509 YASAYR-----GNLSNLATAGLKNRSANPDSDPVENVYFIRAHDSQVQTRIAKLIIRKL 562
Db 779 FVSTYNTVPIDGNAT---TAKVYDETQTSKV-----TYDVNVDEKTIETLFGDN- 826
QY 563 GKTNADGLTNLTDDLNL---KAFDIYNQD-----MNATDKVYYPNNLPMAYAMLMQKQTV 615
Db 827 GKTKNLGVKTTTLTTNANGKATNFTSTDNDALVNAKD-----IAENLNTL 872
QY 616 TRYVYGDMDTDNG-----QYMAKTPFPFNATETLLKGRIKVYAGQAVSYQDWSGIL 669
Db 873 AK-----EIHTTKGTADTALQTFKVKKGATDDETTIVG-----KDGTONGKT 915
QY 670 TSVRVYKGSANSADAGNETRNSGSMALLINNRNPNFRAYNLTLNM-GAAHKQAYRPLLL 728
Db 916 VNTLKLKXGNGLTVA TN---KDGTVTFGINTOSGLKAGDSTTLNKDGLSIKNPASNEQIQ 972
QY 729 STKDGATYLN D-----SDVDSRQYKYVTDQGNLSFSASELQ-----SVANAQVS 773
Db 973 VGADGVKFAKVDKGNSSGTGIDGTSRITKQIGFTGANGSLDTTKPHLTKDKLKVGEVET 1032
QY 774 GMIQVWVPVGAADNQDVRTSPSTQATKD---GNIVHQSDALDSQV---IYEGFSNFQAF 826
Db 1033 ---NTGINAGGKKITNIQSGDITQNSNDNAVTVGRVYDLKTELESKINSAAKTAQNSLHEF 1089
QY 827 AQSPDQ-----YTNVIAKNGDLF-----KSWGITQPEM 855
Db 1090 SVADEQGNHFTVSNPYSSYDTSKTSDVITTFAGENGITTKVNGVVRVIGIDQTKGLTTPKL 1149
QY 856 APO-----YVSSEDG---TPLDSVILN-----GYAFSDRYDLAMSKNNKYGSKQDL 898
Db 1150 TVGNNGNGKIVIDSQKQNTITGLSNTLANVTNDGAGHALS-----QGL 1193
QY 899 ANAIKGLQAGIKVLSDLPVNQLYNIPGK-EVVTATRVNQYGOA---KSGATINKTPYVA 954
Db 1194 ANDTDTRAASI---GDVL-NAGFNLQNGGEAVDF--VSTYDTPDFIDGNATTAKVTY-- 1245
QY 955 NTRSYGDYQEOYGGKFLDDLQKLYPRLF-----STKQJSTGKPIDPSVKITNWSAKYFN 1008
Db 1246 -----DOTSKTSKVYDVNVNVDNKTIETVSDKKL---GVKTTTLTKTSAN 1286
QY 1009 GSNILGRGAKYVLSEGNKYLNADGKLFPTVLNNTYQOPVSANGFISKNGGIHYLDKN 1068
Db 1287 GN-----ATKPSAADGADALVKASD---IATHLNTLAGDIQ-TAKGASQASSASVVDAD 1336
QY 1069 GOBVKRNFKEISGSWYTFSDGKQATGKTKIGNDNTYLFMPNGKQLKEGVWYDQKKAYYVD 1128
Db 1337 GNRV-----IYDS-----TDKTYQVN 1353
QY 1129 DNGRTWTKGFVFRVDDGQKWFYFNGDGTIA-IGLVSLDNRTLYPDAYGYQ----- 1179

Db 1354 DKGOYDKNKEVAKDKLVAQAQ-----TPDGTTLAQMNVKSVINKEQYNDANKKQGINEDNAF 1409
QY 1180 VKG-QTVTINGKSYTFDADQDGLVQTDNANPAPQOAGWKLLGDNQMGVVRKQQLLT--G 1236
Db 1410 IKGLENAAKDKTKNAAVTVGDL-----NA-----VAQTPLTTFAGDTGTTAKKLGETLTIKG 1461
QY 1237 EOTIDGQKVPFQDNGOVQVKGGTATDASGVLRFYDRDQGHQVKGWYSTSDDNWVYVNESG 1296
Db 1462 GGT-DTNKL--TDNNIGVVAG--TDGFTVKLAKDJTNLNSVNAGGTRIDEKGISFVDANG 1516
QY 1297 QVLTGLQITIDGTVTFDDKGIQAKGKAVMDENG-NLRYFD-----ADSGNML- 1342
Db 1517 QAKANTPVLISANGLDLGGKRIISNIGAAVDDNDNAVNFQKQFNEVAKTVNNLNQNSGASLP 1576
QY 1343 -----RDRWKNVDGNWYFNRNGL 1361
Db 1577 FVVTDANGKPINGTDGPKQKAIKGADGKYTHANANGV 1613

RESULT 13
US-11-052-554A-229
; Sequence 229, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; PRIOR FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 229
; LENGTH: 619
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae R6
US-11-052-554A-229

Query Match 3.3%; Score 240.5; DB 7; Length 619;
Best Local Similarity 23.6%; Pred. No. S.1e-07; Indels 91; Gaps 12;
Matches 70; Conservative 42; Mismatches 93;

QY 1066 DKNGQEVKNRFEISGSWYTFSDGKMATGKTKIGNDNTYLFMPNGKQLKEGVWYDQKKAY 1125
Db 394 EKPAAPAKTGWQENGWYFYNTDGSMTG-----WLQN-----NGSW-----Y 432
QY 1126 YYDDNGRTWTKGFVEFRVVDGQDKWRYFNGDGTIAIGLVSLDNRTLYPDAYGVQVKGQTV 1185
Db 433 YLNSNGAMAT--GWLQY-----NGSWYTLNANGAMATGWAKVNGSWTYLNA----- 476
QY 1186 TINGKSYTFDADQDGLVQTDNANPAPQOAGWKLLGDNQMGY-RKDQQLLTGRTQIDGQK 1244
Db 477 --NGAMAT-----GW-LQYNGSWYTLNANGAMATGWAKVNGSW 511
QY 1245 VFFQNGVQVKGGTATDASGVLRFYDRDQGHQVKGWYSTSDDNWVYVNESGQVLTGLQT 1304
Db 512 YVLNANGAM-----ATGWLQY-----NGSWYTLNANGAMATGWAK 546
QY 1305 IDGQTVYFDKGIQAKGKAVMDENGNLRYFDADSGNMLRDRWKNVDGNWYFNRNG 1360
Db 547 VNGSWYTLNANGAMATG---WVKDGTWYLYEASGAMKASQWFKVSDRKWYVYVNLGL 599

RESULT 14
US-10-485-517-141
; Sequence 141, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield

; APPLICANT: Biosynex Incorporated
 ; APPLICANT: Foster, James
 ; APPLICANT: Mond, James
 ; TITLE OF INVENTION: Antigenic Polypeptides
 ; FILE REFERENCE: P100629W0
 ; CURRENT APPLICATION NUMBER: US/10/485,517
 ; CURRENT FILING DATE: 2004-02-02
 ; PRIOR APPLICATION NUMBER: GB 0118825.9
 ; PRIOR FILING DATE: 2001-08-02
 ; PRIOR APPLICATION NUMBER: GB 0200349.9
 ; PRIOR FILING DATE: 2002-01-09
 ; NUMBER OF SEQ ID NOS: 424
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 141
 ; LENGTH: 1290
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus aureus
 US-10-485-517-141

Query Match 3.3%; Score 238.5; DB 6; Length 1290;
 Best Local Similarity 20.0%; Pred. No. 1.9e-06;
 Matches 251; Conservative 155; Mismatches 481; Indels 367; Gaps 59;

QY	40	NNNDGTSVQVNMKVPSPDKPDAQQAQQLQAQAFKAANQADQATATSVSPATDGRVDNQV	99
DB	187	NNVDAAMDKLRQST-ADNATTQKQN-----YTDASQNKQDAYNNAVTTAQQGIIDQTT	238
QY	100	TPAANQPAANVANQDANPATDAGALN-----RQSAADTSTDGKAVPQT-SDQ	146
DB	239	SPTLD-----PVIQAQGVSTTKNALNGNENLEAAKQQAQSQSLGSLDLNNAQKQVTTDQ	295
QY	147	PGHLETVDGKTYVVDANGQBLKNYSMVVDKTYFYFDGQTGEAQTDLPKRTQAQNDNVPDS	206
DB	296	INGAHTVD-EANQIKQNAQNL-NTAM-----GNLQQAIAAD-KDATKATVNFDTA	341
QY	207	YQANNOAYSNEASFEVDNYLTADSHYRPRKILKNGQSWQASSEGDLRPLMTWPDAA	266
DB	342	DQAKQQAINTAVTNAENITISK-----ANGENATQAEVEQAIIQV-----NAA	383
QY	267	TKAAYANF---WAKE---GLISGYRONSANLDAATQNIQSAIEKKIASSEGNTNLWLRDKM	320
DB	384	KQALNGVNAVQHAQDEATALINSNDLNQAQDALKQOVQNAF-----	426
QY	321	SQFVKSQONWSIASNETVYPNQDHMOGGGALLFSNSKOTEHANSWRLLNRNPTFTQKGQ	380
DB	427	-----TVAGVNNVKQTAELNNAVTQAGIADKEQTKADGNGFNADP-----DKQ	472
QY	381	KYFTTNYAGYELLANDVD-----NSNPVQAQBLNHLHYLMNMGDVIWVGDKDAN	430
DB	473	NATVQAQVAKAELISATPDVVVTPSBITAALNKVTOAK--NDLN-----GNTNLATAKQN	525
QY	431	FDGVRVDVNDVNADLLQIORDYKAKYGTDOEKNADHLSILEAWSGNDND-----	483
DB	526	VQHA-IDQLPNLN-----QAQRD-----EYSKQITQATLVPNVNAIQAAATLNDAMTQLKQ	576
QY	484	-----YVKDQNFSLSDNDQSRGMLKAGFYASAYRGNLSNLATAG--LKNRSANPDS	535
DB	577	GIANKAQIKGSYNYH-DADTDKQ-----TAY-----DNAVTKAEELLKQTTNPTWD	621
QY	536	PVNYVPIRAHSEVQTRAKII-----REKLGKTNADGLTNL-TLDDLNKAPD--I	584
DB	622	P-----NTIQOALTQVNDTNOALNGNQKLADAKQDAKTTLTGLTDLHLNDAQKQAL	670
QY	585	YNQDMNATDKVYYPNNLPMAYAMMLNKDVTTRVYVYGDVMTNGQYMAKTFPPYNAIETL	644
DB	671	TTQVEBAPD-IATVNVK-----QNAQNL-----NNAMTNLNAQDKTETLNSINT	717
QY	645	LKGRIKYVAGQAVSKQDWSSGILTSVRYGKGANSASDAGNTETNSGVAL-----LIN	699
DB	718	DADQAKKDAYTNAVSH---AEGILSKA-----NGSNASQTEVEQAMQVRNEAKQALN	766
QY	700	NRPNFRAYRNLTLM--GAAHKSQAVRPLLSTKQGIATVLDNSDVSQYKYKTSQGNL	757

; APPLICANT: Biosynex Incorporated
 ; APPLICANT: Foster, James
 ; APPLICANT: Mond, James
 ; TITLE OF INVENTION: Antigenic Polypeptides
 ; FILE REFERENCE: P100629W0
 ; CURRENT APPLICATION NUMBER: US/10/485,517
 ; CURRENT FILING DATE: 2004-02-02
 ; PRIOR APPLICATION NUMBER: GB 0118825.9
 ; PRIOR FILING DATE: 2001-08-02
 ; PRIOR APPLICATION NUMBER: GB 0200349.9
 ; PRIOR FILING DATE: 2002-01-09
 ; NUMBER OF SEQ ID NOS: 424
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 141
 ; LENGTH: 1290
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus aureus
 US-10-485-517-141

Query Match 3.3%; Score 238.5; DB 6; Length 1290;
 Best Local Similarity 20.0%; Pred. No. 1.9e-06;
 Matches 251; Conservative 155; Mismatches 481; Indels 367; Gaps 59;

QY	40	NNNDGTSVQVNMKVPSPDKPDAQQAQQLQAQAFKAANQADQATATSVSPATDGRVDNQV	99
DB	187	NNVDAAMDKLRQST-ADNATTQKQN-----YTDASQNKQDAYNNAVTTAQQGIIDQTT	238
QY	100	TPAANQPAANVANQDANPATDAGALN-----RQSAADTSTDGKAVPQT-SDQ	146
DB	239	SPTLD-----PVIQAQGVSTTKNALNGNENLEAAKQQAQSQSLGSLDLNNAQKQVTTDQ	295
QY	147	PGHLETVDGKTYVVDANGQBLKNYSMVVDKTYFYFDGQTGEAQTDLPKRTQAQNDNVPDS	206
DB	296	INGAHTVD-EANQIKQNAQNL-NTAM-----GNLQAQIAD-KDATKATVNFDTA	341
QY	207	YQANNOAYSNEASFEVDNYLTADSHYRPRKILKNGQSWQASSEGDLRPLMTWPDAA	266
DB	342	DQAKQAQYNTAVTNAENITISK-----ANGENATQAEVEQAQIKQV-----NAA	383
QY	267	TKAAYANF---WAKE---GLISGYRONSANLDAATQNIQSAIEKKIASSEGNTNLWLRDKM	320
DB	384	KQALNGVNAVQHAQDEATALINSNDLNQAQDALKQOVQAT-----	426
QY	321	SQFVKSQONWSIASENETVYPNQDHQGGALLFSNSKOTEHANSWRLLNRNPTFTQKGQ	380
DB	427	-----TVAGVNVNKQTAELNNAVMTQLAQGIADKEQTKADGNGFNADP-----DKQ	472
QY	381	KYFTTNYAGYELLANDVD-----NSNPVQAEQLNHLHYLMNKGDIIVMGDKDAN	430
DB	473	NATVQAQVAKAELISATPDVVVTPSBITALNKNVTOAK--NDLN-----GNTNLATAKQN	525
QY	431	FDGVRVDVNDVNADLLQIORDYKAKYGTDOEKNADHLSILEAWSGNDND-----	483
DB	526	VQHA-IDQLPNLN-----QAQRD-----EYSKQITQATLVPNVNAIQAAATLNDAMTQLKQ	576
QY	484	-----YVKDQNFSLSDNDQSRGMLKAGFYASAYRGNLSNLATAG--LKNRSANPDS	535
DB	577	GIANKAQIKGSYNYH-DADTDKQ-----TAY-----DNAVTKAEELLKQTTNPTWD	621
QY	536	PVNYVYFIRAHSEVQTRAKII-----REKLGKTNADGLTNL-TLDDLNKAPD--I	584
DB	622	P-----NTIQOALTQVNDTNOALNGNQKADAKQDARTLTGLTDLHLNDAQKQAL	670
QY	585	YNQDMNATDKVYYPNNLPMAYAMMLNKDVTTRVYVYGDVMTNGQYMATKTPFYNAIETL	644
DB	671	TTQVEBAPD-IATVNVK-----QNAQNL-----NNAMTNLNAQDKTETLNSINTP	717
QY	645	LKGRIKYVAGQAVSKQDWSSGILTSVRYGKGANSASDAGNTETNSGVAL-----LTN	699
DB	718	DADQAKKDAYTNAVSH---AEGILSKA-----NGSNASQTEVEQAMQVRNEAKQALN	766
QY	700	NRPNFRAYRNLTLM--GAAHKSQAVRPLLSTKQGIATVLDNSDVDSRQYKYTQSQNL	757

Db 617 TDAVTKQLKDAKPTLNAGDGIS--INSNNGDLVDSSGNIITPTTNYISVKT-----TKLNS 670
Qy 195 TQOANQDNVPDSYQANNOAYSNEASFETVDNYL-----TADSWYRPRKILKNQOSWQAS 249
Db 671 NGTSGN---NKFSVSNADHNSLVTAKDLADYLNKVNETADALSALPSFKVQNGDSSNNAI 726
Qy 250 SEGDLRPLMTWPDAAATKAAAYANFWAKEGLISGSYRQNSANI--DAATQNIQSAIEKK- 306
Db 727 TVG-----KDTNGKTFNTLKLKG-----ENGVNITTNRAITGTFTFGIDQSN 767
Qy 307 -----IASBGTNWLNRDKMSQFVKSONOWSIASENETVYPNQDHHMOGGALLFNS-- 356
Db 768 GLITPKLTVGSTNGNRLVIEQVPSADGNSTKNIILKGLSPTLPSIASPSGRNIALGNTIE 827
Qy 357 -KDETHANSDBWLNRNPTFO-TGKQKYFTTNYAGYELLANDVNSNPNVQOAEQLHLH 414
Db 828 EKDKSNAASIDVDVLNAGFNKNGKDKDFVSTYDVTDFIDGNATTATVYDEANQTSKVA 887
Qy 415 YLMNMGDIWM---GDKDANFDGVR---VDAVDNVNADLLQIORDYYKAK----- 457
Db 888 YDYNVDEKTIELTGDNGKKQLGVKTIKLTETSTNGNATFTSTDDHALVKASDIAGNLMT 947
Qy 458 -----YGT-----DQNEKNAIDHLSIL---EAWSGNDNDY-VKQDNFNS 492
Db 948 LAEEIHTTKGTANTALQTTVKKVBENDK-ADDTNATVKGKGTSGKVNTLKKGKGLD 1006
Qy 493 LSIQND-----QRSGMLKAFGYASAYRGNLSNLATAGL--KNRSANPD-----SDPVP 538
Db 1007 IKTKDKGTFTFGINTQSG-LKA-----GDSFTLNNGLSIKNTASNEQIQVGADGVK 1057
Qy 539 NVVFRADHSEVQTRI---AKIIREKLGKTNADGLTNLTDLNLKAFDIYNQDMNATDKV 595
Db 1058 ---FAMVNGVVGAGIDGTRITRDEIGFTGTNGSLDKSKPHLSK-----DCINAGGK- 1107
Qy 596 YYPNNLPMAYANMLQKDTVRVYVYGDVMTDNGQYMATKTPFYNAIETLLKGRIKYVAGG 655
Db 1108 -----KITNIQSGE-----IAKNSHDAVTGG 1128
Qy 656 QAVSYNQDWSGILTSVRYKGA-----NSASDAGNTETRNSGMALLNNRPN----FRAY 707
Db 1129 KIYDLKTELENKISSAKTAQNSLHEFSVADEQGNFTVSNPYSSYDTSKTSDDVITFAGE 1188
Qy 708 RMLT--LWNGAAH-----KSOAYRPLLLSTDKGTATYLANDSDVRSORYKYTDSQNLSP 759
Db 1189 NGITTKVNRGVVRVGDIDQTKGLTTPKLTGVNNGNGGIVINSQ-----NGQNTITG 1238
Qy 760 SASELQSVANAQVSGMIQVWVPVGAADNQDVRTSPSTQATKDCGNIYHQSD-----ALDSQV 815
Db 1239 LSNLTANVTN-----DKGSVRT-----TEQGNIIKDEKTRAASIVDV 1276
Qy 816 IYEGFSNFOAQAQSPDQYT-----NAVIK--NGDLFKSWGJ-----TQFEM 855
Db 1277 LSAGF-NLQNGEAVDFVSTYDTPNFANGNTTTAKVYDDTSTKTSKVYDVNVVDDTTIEV 1335
Qy 856 AFQYVS-----SEDTGLDVLINGYAFSDRY-----DLAMSKNKYKSKQDLANAIGL 905
Db 1336 KDKKLGKVTTLTSTGTGANKFALSNOATGDALVKAASDIVAHLNLTLSGDIQTAKGASQAN 1395
Qy 906 QSAGIKVLSDLVPNQLNPLGKEVVTATRVNOYGOAKSGATINKTPYVANTRSYGDYQEQ 965
Db 1396 NSAG-----YVDAGNKVIYDSTDNKYIQANNDGTVDKTEKVAKDKLVAQAQTP 1444
Qy 966 YGKGFLLDLQKLPRLFSTKQISTGKPIDPSVKITNWSAKYFNNGSILGRGAKYVLSGN 1025
Db 1445 DG-----TLAQM-----NVKSVINKEQVNDANK-----KQGINEDNAFVKGLEKAASD-N 1488
Qy 1026 KYLNADGKLPFLPTV--LNNYGOPOQVSA--NGFISKNGGIHYLDKNGQEVKNRFXE--- 1078
Db 1489 KTKNAA-----VTVGDLNVAQTPLTFAAGDTGTGTAKLGETLTIKGGQDTNKLTDNNI 1542
Qy 1079 --ISGSWYVDSG---KNATGKTKIGNDTYLPMFMPNGKQKKEGVMYDGKKAYYYDDNGRT 1133
Db 1543 GVVAG-----TDGFTVKLAKDLTNLSVN-----AGGTKIDEXISFVDANGQA 1586

Qy 1134 WTNKGFVEFRVDGQDKWRYFNGDGTIAIGLYSLDNRLTYFDAYGYQVKGTQVTINKKSYT 1193
Db 1587 KAN-----TPVLSANGLDLGGKVISNVGKG-T 1612
Qy 1194 FDADQGDLLVQTDNANPAPQGOAGWKLLG---DNQWGYR-----K 1229
Db 1613 KDTDAANVQQLNEVR-----NLLGLGNDNADGNQVNIADIKKDPNSGSSSNRTVIK 1663
Qy 1230 DQQLLTGSGQTTIDGQKVFQDNGVQV---KGGTAT--DASGVLRFYDRDQGHQVKGWYSTS 1285
Db 1664 AGTVLGGKGNNDTEKL--ATGGVQGVVDKDGANGDLSNV-----WVKTQ 1706
Qy 1286 DDNWWYVNESQVLTGLQTFIDGQTYVFDKGIQAKGKAVWDENGNLRYFDADSGN 1340
Db 1707 KDG-----SKALLATYNAAGQTNVYVNTNPPABAIDRI--NEQG-IRFPFHVNDGN 1752

Search completed: February 11, 2006, 20:58:50
Job time : 18.7677 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 19:15:50 ; Search time 28.1742 Seconds
(without alignments)
4661.567 Million cell updates/sec

Title: US-10-797-821-39
Perfect score: 7230
Sequence: 1 MEKNRYKLHKVKQWVAIG.....WKNVDGNWYFNRNGLATRW 1365
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	7230	100.0	1365	2 A41483	glucosyltransferas
2	3174	43.9	1431	2 A45866	dextranucrase (EC
3	3079.5	42.6	1449	2 T30857	glucosyltransferas
4	3076.5	42.6	1449	2 T30552	glucosyltransferas
5	3026.5	41.9	1577	2 T30858	glucosyltransferas
6	3023	41.8	1592	2 A38175	glucosyltransferas
7	2945.5	40.7	1475	2 B33135	gtfB protein precu
8	2893.5	40.0	1375	2 JT0345	dextranucrase (EC
9	2863.5	39.6	1599	2 S22737	glucosyltransferas
10	2804	38.8	1518	2 A44811	glucosyltransferas
11	2702	37.4	1508	2 T31098	probable dextranu
12	2593.5	35.9	1290	2 JCS473	dextranucrase (EC
13	543	7.5	2817	2 B97033	unchacterized pr
14	367.5	5.1	2710	2 A37052	toxin A - Clostrid
15	358.5	5.0	2364	2 I40884	cytotoxin L - Clos
16	334.5	4.6	2367	2 S70172	toxin B - Clostrid
17	327.5	4.5	2366	2 S10317	toxin B - Clostrid
18	323	4.5	2178	2 S55805	alpha-toxin - Clos
19	316.5	4.4	563	2 A37184	glucan-binding pro
20	306.5	4.2	2334	2 S12920	cell wall-associat
21	285	3.9	648	2 S10869	enterotoxin A - Cl
22	273	3.8	1441	2 B86807	hypothetical prote
23	270.5	3.7	690	2 F98114	choline-binding pr
24	267	3.7	2167	2 AF1489	cell wall-associat
25	260	3.6	1999	2 AB2018	hypothetical prote
26	258.5	3.6	1385	2 D89824	hypothetical prote
27	256	3.5	1463	2 T30290	AAS surface protei
28	247.5	3.4	2703	2 H81193	hemagglutinin/hemo
29	247	3.4	583	2 S57721	capB protein - Clo

ALIGNMENTS

RESULT 1

A41483
glucosyltransferase (EC 2.4.1.-) gtfS precursor - Streptococcus sobrinus
C:Species: Streptococcus sobrinus
C>Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 15-Mar-2004
C:Accession: A41483
R:Gilmore, K.S.; Russell, R.R.B.; Ferretti, J.J.
Infect. Immun. 58, 2452-2458, 1990
A:Title: Analysis of the Streptococcus downei gtfS gene, which specifies a glucosyltran
A:Reference number: A41483; MUID:90316665; PMID:2142479
A:Accession: A41483
A:Molecule type: DNA
A:Residues: 1-1365 <GIL>
A:Cross-references: UNIPARC:UPI000012BCB6; GB:M30943; NID:g153652; PIDN:AAA26898.1; PID
C:Genetics:
A:Gene: gtfS
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 100.0%; Score 7230; DB 2; Length 1365;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEKNRYKLHKVKQWVAIGVTTVTLSTFLAGQVVAADTNNNDGTSVQVKNKVPSPKFD 60
DB 1 MEKNRYKLHKVKQWVAIGVTTVTLSTFLAGQVVAADTNNNDGTSVQVKNKVPSPKFD 60

QY 61 AQAQNGQLAQAAMFKAANOADQATTSQVSPATDGRVDNQVTPAANQPAANVANQDVANPAT 120
DB 61 AQAQNGQLAQAAMFKAANOADQATTSQVSPATDGRVDNQVTPAANQPAANVANQDVANPAT 120

QY 121 DAGALNROSAADTSTDGKAVPQTSDPQGHLETVDGKTYVDGQRLKNYSVIDGKTTY 180
DB 121 DAGALNROSAADTSTDGKAVPQTSDPQGHLETVDGKTYVDGQRLKNYSVIDGKTTY 180

QY 181 FDQGTGEAQTDLPKTQANQDNVPSYQANNQAYNEASSFETVDNYLTADSWYPRKIL 240
DB 181 FDQGTGEAQTDLPKTQANQDNVPSYQANNQAYNEASSFETVDNYLTADSWYPRKIL 240

QY 241 KNGSQWQASSEGDLRPIILMTWPDATKAAVANFPAWKEGLISGSYRQNSANLDAATQNIQ 300
DB 241 KNGSQWQASSEGDLRPIILMTWPDATKAAVANFPAWKEGLISGSYRQNSANLDAATQNIQ 300

QY 301 SAIKKIASEGNTWLRDKMSQFVKSONQWSTASENETVYPNODHMQGALLFNSKQTE 360
DB 301 SAIKKIASEGNTWLRDKMSQFVKSONQWSTASENETVYPNODHMQGALLFNSKQTE 360

QY 361 HANSDWRLNRPFTQTKQKYFTTNAYGYELLANDVNSNPVVOAEQNLHLYMNGW 420
DB 361 HANSDWRLNRPFTQTKQKYFTTNAYGYELLANDVNSNPVVOAEQNLHLYMNGW 420

QY 421 DIVMGDKDANFDGVRVDVNVNADLLQIQRDYKAKYGTQDQNEKNAIDHLISLEAWSGN 480

hypothetical prote
cspB protein - Clo
choline binding pr
fibrinogen-binding
surface protein ps
surface protein ps
hypothetical prote
dextranase inhibito
hypothetical prote
pullulanase (EC 3.
hypothetical prote
probable peptidogl
pneumococcal surfa
hypothetical prote
toxin-like outer m
choline binding pr

Db 421 DIVMGDKANFDGVRVDAVDNVNADLLQIQDYKAKYCTDQNEKNAIDHLSILEAWSN 480
QY 481 DNDYVKQNNFSLSDNDORSGLMKAFGYASAYRGNLSNLTATAGLKNRSANPDSPVPNY 540
Db 481 DNDYVKQNNFSLSDNDORSGLMKAFGYASAYRGNLSNLTATAGLKNRSANPDSPVPNY 540
QY 541 VFIRAHDSVOTRIAKIREKLGKTNADGLTNLTDDLNKAFDIYNQDMNATDKVYYPNN 600
Db 541 VFIRAHDSVOTRIAKIREKLGKTNADGLTNLTDDLNKAFDIYNQDMNATDKVYYPNN 600
QY 601 LPMAYAWMLQNKDVTTRVYGYDMYTDNGQYMATKTPFYNAIETLLKGRIKYYVAGQAVSY 660
Db 601 LPMAYAWMLQNKDVTTRVYGYDMYTDNGQYMATKTPFYNAIETLLKGRIKYYVAGQAVSY 660
QY 661 KQDWSSGILTSVRYKGGANSASDAGNTETRNSGMALLINRPNFRAYRNLTLNMGAAHKS 720
Db 661 KQDWSSGILTSVRYKGGANSASDAGNTETRNSGMALLINRPNFRAYRNLTLNMGAAHKS 720
QY 721 QAYRPLLSTKOGIATYLNDSVDVSRQYKTYDSQGNLSPSASELOSANAOVSGMIQVWV 780
Db 721 QAYRPLLSTKOGIATYLNDSVDVSRQYKTYDSQGNLSPSASELOSANAOVSGMIQVWV 780
QY 781 PVGAADNQDVRTSPSTQATKDGNIIHQSDALDSQVIYEGFSNFQAFQSPDOYTNAVIAK 840
Db 781 PVGAADNQDVRTSPSTQATKDGNIIHQSDALDSQVIYEGFSNFQAFQSPDOYTNAVIAK 840
QY 841 NGDLFKSWGITQFEMAPQVVSSEDTGFLDSVILNGYAFSDRYDLAMSKNKGYSKQDLAN 900
Db 841 NGDLFKSWGITQFEMAPQVVSSEDTGFLDSVILNGYAFSDRYDLAMSKNKGYSKQDLAN 900
QY 901 AIKGLQSGIKVLSLDPNQLNLPKGVVTVATRVNOYGOAKSGATINKTPVANTRSYG 960
Db 901 AIKGLQSGIKVLSLDPNQLNLPKGVVTVATRVNOYGOAKSGATINKTPVANTRSYG 960
QY 961 DYQEQYGGKFLDQKLYPRLFTSKQISTGKPIDSPSVKITNWSAKYFNGSNILGRGAKYV 1020
Db 961 DYQEQYGGKFLDQKLYPRLFTSKQISTGKPIDSPSVKITNWSAKYFNGSNILGRGAKYV 1020
QY 1021 LSEGKNYMLADGKFLPVLVNTYGPQVVSANGFISKNGGTHYLDKNGQEVKNRPFKEIS 1080
Db 1021 LSEGKNYMLADGKFLPVLVNTYGPQVVSANGFISKNGGTHYLDKNGQEVKNRPFKEIS 1080
QY 1081 GSWYVFDSDGKMATGKTIGNDTYLFPMPNGKQKGVWYDGGKAYYYDDNGRTWTKGPFV 1140
Db 1081 GSWYVFDSDGKMATGKTIGNDTYLFPMPNGKQKGVWYDGGKAYYYDDNGRTWTKGPFV 1140
QY 1141 EFRVDQDKWRYFNGDGTIAIGLVSLDNRTLYFDAYGYQVKGQTVTINGKSYTFDADQGD 1200
Db 1141 EFRVDQDKWRYFNGDGTIAIGLVSLDNRTLYFDAYGYQVKGQTVTINGKSYTFDADQGD 1200
QY 1201 LVQTDNANPAQOQAGKLLGDNQNGYRKDGQLLTGEQTI DGQVFFQDNGVQVKGGTAT 1260
Db 1201 LVQTDNANPAQOQAGKLLGDNQNGYRKDGQLLTGEQTI DGQVFFQDNGVQVKGGTAT 1260
QY 1261 DASGVLFYDRDQGHQVKGWYSTDNDNMYVNESQVLTGLQTDGQTVYFDDKGIQAK 1320
Db 1261 DASGVLFYDRDQGHQVKGWYSTDNDNMYVNESQVLTGLQTDGQTVYFDDKGIQAK 1320
QY 1321 GRAVWDENGLRYFDADSGNMLRDRKNDVNGWYFNRRNGLATRW 1365
Db 1321 GRAVWDENGLRYFDADSGNMLRDRKNDVNGWYFNRRNGLATRW 1365

RESULT 2
A45866
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans
C:Species: Streptococcus mutans
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Mar-2004
C:Accession: A45866
R:Honda, O.; Kato, C.; Kuramitsu, H.K.
J. Gen. Microbiol. 136, 2099-2105, 1990
A:Title: Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the glucosyl
A:Reference number: A45866, MUID:91100958, PMID:2148600

A:Accession: A45866
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1431 <HON>
A:Cross-references: UNIPARC:UPI000017AC5C; GB:M29296
C:Keywords: glycosyltransferase; hexosyltransferase
F:181-201/Domain: cpl repeat homology <CP1>
F:1127-1146/Domain: cpl repeat homology <CP2>
F:1192-1211/Domain: cpl repeat homology <CP3>
F:1257-1276/Domain: cpl repeat homology <CP4>
F:1277-1297/Domain: cpl repeat homology <CP5>
F:1321-1340/Domain: cpl repeat homology <CP8>
F:1341-1361/Domain: cpl repeat homology <CP6>
F:1385-1404/Domain: cpl repeat homology <CP7>

Query Match 43.9%; Score 3174; DB 2; Length 1431;
Best Local Similarity 47.3%; Pred. No. 2.4e-153;
Matches 677; Conservative 213; Mismatches 436; Indels 104; Gaps 38;

QY 1 MEKNRYKLVKVKKQWVAIGVTT--VTLSFLAGGQVVAADTNNDGTSV-----QVNMKV 53
Db 1 METKRYKQVHKVHKHVVAVASGLITLTGTTGLSSVSAETEQQTSDKVVTQKSDDKAA 60

QY 54 PSDPKFDA---QAQNGQLAQAAMFKAANOADOTATSOVSPATDGRVDNQVTPAANOAA 109
Db 61 SESSQTDAPKTKAQTEQ-TQAQ-SQANVAD-TSTS-----ITKETPSQNIITQANSDDKT 113

QY 110 VANQDVANPATDAGALNRQSAADTSTDGKAVP-----QTSDDPG-----H 149
Db 114 VTNKSEEAQTSSERTKQSEEAQTASSQAULTQAQAEITKQRTAAQENKNPVDLAAIPN 173

QY 150 LETVDGKTYVDANGQRLKNYSWIDGTYTFDGTGEAQTDLP----KTGAQNDVDP 205
Db 174 VKQIDGKTYTGSQDQPKKFNALTVNNKVLVPDKNTG-ALTDTSYQFQKQGLTKLNN--- 229

QY 206 SYQANNOAYSNEASFETVDNYLTADSWYRPRKILKNGOSWASSEGDLRPLIMTWPPDA 265
Db 230 DYTPHQVQINVENTSLETIDNYVADSWYRPRKILKNGKTWTASSEDLRPLIMSWWPK 289

QY 266 ATKAAAYANFWAKEGLISG-SYRQNSA--NLDAATQNIQSAIEKKIASEGNTWLRDKMSQ 322
Db 290 QTOIAYLVNMQOGLGTGENYTDSSQESLNLAQTQVQVKEIKISQTOQTQWLRLDI 349

QY 323 FVKSQNQSIASENETVYPNQDHQMGALLPSNSKQDTEHANSQDRLLENRPFTQTKQKY 382
Db 350 FYKTOPNWSQTESDTSAGEKHLQGLGALLYSNSDKTAYANSYDLLNARTPTSTQTKPKY 409

QY 383 FTTNYA-GVELLLANDVNSNPVQAEQLNHLHYLMNWGDI VMGDKDANFDGVRVDAVDN 441
Db 410 FEDNSSGGYDFLLANDIINSPVQAEQLNHLHYLMNYGSIIVANDPEANFDGVRVDAVDN 469

QY 442 VNADLLQIQDYKAKYGTQNEKNVAIDHLSILEAWSNDNDYVKQNNFSLSDNDORS 501
Db 470 VNADLLQIASDYLKAHYGVDKSEKNAINHLSILEAWSNDNDPOYNKDTKGAQLPIDNKLRL 529

QY 502 GMLKAF-----GYASAYRGNLSNLTATAGLKNSA-NPQSDPVPNYVFIHAKDSEVQT 552
Db 530 SLLYALTPLPEKQASNKNEIRSGLEPVITNSLNNSASGKNSERWANYIFIHAKDSEVQT 589

QY 553 RIAKIIREKLGKTNADGLTNLTDDLNKAFDIYNQDMNATDKVYYPNNLPMAYAWMLQNK 612
Db 590 VIAKIICAQI-NPKTDGLT-FTLDELKQAFKIYNEQMRQAKKQYQTSNIPTAYALMLSNK 647

QY 613 DTVTRVYGYDMYTDNGQYMATKTPFYNAIETLLKGRIKYYVAGQ--AVSYKQ-----DW 664
Db 648 DSITRLYYGYDMYSDGQYMATKSPYYDAIDTLLKARIKYAAGGQDMKITYVEGDKSHMDW 707

QY 665 S-SGLITSVRVYKGGANSASDAGNTRNSGMALLINRPNFRAYRN--LTLNMGAAHKSQ 721
Db 708 DYTGLVTSVRYGTGANEATDQGEATKQGMVAVITSNPNPSLKNQNDKVINVMGAAHKNQ 767

QY 722 AYRPLLSTKOGIATYLNDSVDVSRQYKTYDSQGNLSPSASELOSANAOVSGMIQVWV 781
Db 781 AYRPLLSTKOGIATYLNDSVDVSRQYKTYDSQGNLSPSASELOSANAOVSGMIQVWV 781

```
Db 768 EYRPLLLTTKDGLTSTSDAAKS-LYRKTNDKGELVFDASDIQGYLNPQVSGYLAVWP 826
Qy 782 VGAADNQDVRTSPSTQATKGNIIYHOSDALDSQVIYEGFSNFQAFQSPDQYTNVIAKN 841
Db 827 VGASDQDVRVAASNAKANATQGVYESSALDSQLIYEGFSNFQDFVTKDSYTNKKIAQN 886
Qy 842 GDLFKSWGITQFEMAPQYVSSEDTGLDLSVLNGYAFSDRYDLAMSKNNKYKSGQDLANA 901
Db 887 VOLFKSWGITSFEMAPQYVSSEDSGLDSIIQNGYAFEDRYDLAMSKNNKYKSGQDLANA 946
Qy 902 IKGLQAGKIVLSDLPNOLYNLPKGEVVTATRVNOYGOAKSGATINKTPYVANTRSYG- 960
Db 947 VKALHSGSQIADWPDQIYNLPKGEVVTATRVNDYGEYKDSSEIKNTLYAANTKNGK 1006
Qy 961 DYQEQYGGKPLDQLKLPRLFTSKQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKYV 1020
Db 1007 DYQAKYGGAPLSLAAYKPSIFNRTQISNGKIDPSEKITAWKAKYFNGTNILGRGVYV 1066
Qy 1021 LSE--GNKYNLADGKLPFTVLNNTYGGQVQVSANGFISKNGGIIHYLDKNGQEVKRF-K 1077
Db 1067 LKDNASDKYFELKGNQTYLPKQMTN-----KEASTGFVNDGNGMTFYSTSGYQAKNSFVQ 1121
Qy 1078 EISGSWYFSDSGKMATKTKIGNDTYLFPNGKQLKEGVW--YDCKKAYYYDDNGRTWT 1135
Db 1122 DAKGNWYFDNNGHMYVGLQNGEVQYFLSNGVQLRESFLENADGSK-NYFGLGNRYS 1180
Qy 1136 NKGFEVFRYDGDQKRYFNGDGTIAIGVSLDNRITLYFDAYGYQVKGQTVT-INGKSYTF 1194
Db 1181 N-GYYSF--DNDSKRYFPDASGVMAVGLKTINGNTQYFQDQGYQVKGAMITGSDGKKRYF 1237
Qy 1195 DADQGLDVTDNANPAPOAGKWLGDNQWGY-RKDGQLLTGEQITDQKVPFQDNGVQ 1253
Db 1238 DPGSGNMAVNFANDK-----NGDWYLYNSDGIALGVQVQINGKTYFFGQDGKQ 1286
Qy 1254 VKGGTATDASGVLYRFDYRQGHQVKGWYSTDDNWWYVNESQVLTGLQITDQVTFD 1313
Db 1287 IKGKIITD-NGKLIYFLANGELARNIFATDSNNWYFPGSDGVAVTGSQITAGKLYFA 1345
Qy 1314 DRGIOAKGKAVDENGRLRYFDADSGNMLRDRWK-NVDGNWYFNRNGLA 1362
Db 1346 SDGKQVKGSGFV-TYNGKVHYHADSGELQVNRFEADKGNWYLYDSNGEA 1394

RESULT 3
T30857
Glucosyltransferase - Streptococcus salivarius
C.Species: Streptococcus salivarius
C.Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C.Accession: T30857
R.Simpson, C.L.; Giffard, P.M.; Jacques, N.A.
Infect. Immun. 63, 609-621, 1995
A.Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for pri
A.Reference number: Z20909, MUID:95122197, PMID:7822030
A.Accession: T30857
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-1449 <SIM>
A.Cross-references: UNIPROT:Q55264; UNIPARC:UPI00000B166E; EMBL:L35495; NID:g662378; PID
C.Genetics:
A.Gene: gtfI

Query Match 42.6%; Score 3079.5; DB 2; Length 1449;
Best Local Similarity 45.0%; Pred. No. 1.6e-148;
Matches 667; Conservative 206; Mismatches 451; Indels 157; Gaps 35;

Qy 1 MEKNRLYKLVKKQKQWVIAVTTVLISFLAGGVAAADNNNDGTSVQVKNWVPSDPKPD 60
Db 1 MDKRVHYKMKVKKQWVTVIAVTLGLSL-----GAVSAVSLGTNDG-----VVQADEHTD 48
Qy 61 AQAOQGLAQAFKAANO--ADQATATSVSPATDGRVDNQVTPAA-----NOPAAN 109
Db 49 ATVAIPDITVDRTGTVNSDNTAAQDPPTTAAATNDVATD-QATPTATFDLTDTTNTVAAN 107
```

```
Qy 110 VANQDVANPATDAGALNRQSA-----DTSTDGKAV----- 140
Db 108 -AVDTVATVGTDRAAATTNDTTATNDTAVDTTNNNTTDTTVDTRAATTERRATGARRGP 166
Qy 141 -----POT-----SDQPHLETVDGKTYVVDANGQBLKN 169
Db 167 TGRRRATPVNGTNNANNTVTVVNNDLPATNNVVTDPGSHIKTINGKQYVVEDGTIRKN 226
Qy 170 YSNVIDGKTYVDPGTGEAQT-----LPKTGOANQDNVDPVQAN-----NOAYSNEA 218
Db 227 YVLERIGGSOYFNAETGELSNQKRYFDKNGGTGSS--ADSTNTVTVNGCDKNAFYGTTD 284
Qy 219 SPSFETVNYLTADSWTRPKILKNGSOWASSEGDLRPIILMTWMPDAATAKAAAYANFAWE 278
Db 285 KDIELVDGYFTANTWYRPEIKLKGKEWTASTENDKRPILLTVWVPSKALQASLYNTMKEQ 344
Qy 279 GL-IGSVYRQNSA--NLDAATQNIQSAIEKKIASEGNTNMLRDKMSQFVKSONQWIAS 335
Db 345 GLGTNQTYTSFSQTMQDALEVKRIEERAREGNTDMLRTTIKNFKVKTQPGMNSTSE 404
Qy 336 NETVTPNODHMOGGALLFNSKDTETHANSDWLLNRNPTFQTKQ--KYP-TTNYAGYEL 392
Db 405 N--LDNNDHLOGGALLYNNDSRTSHANSDYRLNARTPTSQGKNPKYTKDTSNGGFEP 461
Qy 393 LLANDVNSNPVVAEQALNHLHYLMNWGDIVMGDKDANFDGVRVADVAVNNADLLQIQRD 452
Db 462 LLANDIDNSNPAVQAEQNLWLVHINIGTITGSEDEDFGVRVADVAVNNADLLQIASD 521
Qy 453 YKAKYGTQDNEKNAIDHLSIIEANSGNDNDYVKQDNFSLSDNDORSGLKAFGYASA 512
Db 522 YPKAKYGADQSDQAKHLSIIEANSHNDAYYNEDTKGAQLPMDPMHLALVYLLRPIG 581
Qy 513 YEGNLSNLATAGLKNRS-ANPDSDPVNPVFIKRAHSEVQTRIAKIREKLGKTNADGLT 571
Db 582 NRSQVEPLISNLSNDRSESGKNSKRWYAFVRAHDSQVOSTIGQIKNEINPQSTG--N 639
Qy 572 NUTLDDLKAFDIYNODMNATDKVYYPNNLPMAWMLQNKQDVTTRVYVYGDYTDNGQYM 631
Db 640 TPTLDEMKAPEIYNKDMRSANKQVYNIPSAYALMLTHKDTVPVRYVYGDYTDGQYM 699
Qy 632 AKTTFPNAIETLLKRIKYVAGGQ--AVSY-----KQDW-SSGILTSVRYKGANASD 683
Db 700 AQKSPYDAIETLLKGRIRYAAAGGQDMKVNYTGYGNTNGMDAAGVLTSVRYGTGANASD 759
Qy 684 AGNTETRNSGMALLNNRPNFRAYRNLTLNMGAAKHSQAYRPLLSTKDGATYLNDSV 743
Db 760 TGTATRNQGMAYIVSNQPALRLTSLNLTINMGAAHNRQAYRPELLTTNDGVATYLNDSA 819
Qy 744 DSRQKYTSDSQNLFSASELOSANAOVSGMIQVWVPVGAADNQDVRTSPSTQATKGN 803
Db 820 NG-IVKYTDGNGNLTFSEANEIRGNPQVDGYLAVWVPVGAASENQDVRVAPSKENSSGL 878
Qy 804 IYHQSADLDSQVIYEGFSNFQAFQSPDQYTNVATKNGDLPKSWGITQFEMAPQVVSSE 863
Db 879 VYESNAALDSQVIYEGFSNFQFVQNPQSYTNKKIAENANLFSKSWGITSFEPAPQVSSD 938
Qy 864 DGTFLDSLVLNGYAFSDRYDLAMSKNNKYKSGQDLANAIGKQAGIKVLSLDPVNLQYN 923
Db 939 DGSFLDSVLQNGYAFTRDYDGMKDNKNGSLADLKAALKSLHAGVISAIDAVWPQIYN 998
Qy 924 LFGKEVVTATRVNOYGOAKSGATINKTPYVANTRSYG-DYQEQYGGKFLDLDLQKLYPRL 982
Db 999 LFGDEVVTATRVNNYGETKQGAIDHSLYAAKTRTFGNDYQGYGAGFLDELKRLYPQIF 1058
Qy 983 STKQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKYVLSSEG-NKYNLADGKLPFTVL 1041
Db 1059 DRVQISTGKRMITDEKITQWSAKYMNMTNILDGRSEYVLKNGLNGYGTNGGKVSLPKV 1118
Qy 1042 NNTYQGPQVPSANG-----FISKNGGIIHYLDKNGQEVKQRF-KEISGSWYVYFSDGKMWG 1095
Db 1119 GSNQSTGNDGNDGSGKPEKFLFSVRYRNNNGYAKNAFINKDNGNVYFDSNGMAVG 1178
Qy 1096 KTKIGNDYLFPNGKQLKEGVWYDQK-KAYYYDDNGRTWTN-----KGFEV 1141
```

Db 1179 EKTIDGKQYFFLANGVLEDDGYRQNRGQVFFYDQGVNLNANGKODPKPDNNNNASGRNQ 1238
QY 1142 FRVDGQDKWRYFNGDGTIAIGLVSLDNRITLYFDAYGVYQVKGQTVTINGKSYTFDADQDGL 1201
Db 1239 FVQIGNVWAYVDGNGKRVTHQNINGQBELFFDNNNGVQVKGRTVNGGAIRYYDANSSEM 1298
QY 1202 VOTDANPAQOAGKWLGDNQWGY-RKDGQLLTGEGTIDGQVFFQDNGVGVKGGTAT 1260
Db 1299 ARNRFAETEP-----GVWAYFNNDGTAVKGSQNINGQDLYFDQNGRQVKGALA- 1346
QY 1261 DASGVLYRFDYDRDQGHQVKGWYSTSDNNWVYVNESGOVLTLGTLTIDGQTVYFDDKGIQAK 1320
Db 1347 NVDGNLRYDVNSG-ELYRNRPHEDGSMYFDGNGVAKGMVNINGQNLFFDNNKGQIK 1405
QY 1321 GKAVWDEGNLRYFDADSGNMLDRWKNVD-GNYYFNRNG 1360
Db 1406 GHLV-RVNGVRYFDPNSGEMAVNRWVEVSPGWVYFDGEG 1445

RESULT 4
T30552
glucosyltransferase N - Streptococcus salivarius (fragment)
C:Species: Streptococcus salivarius
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30552
R:Jaffe, R.I.
submitted to the EMBL Data Library, February 1998
A:Description: Streptococcus salivarius V1477 gtfN.
A:Reference number: Z20854
A:Accession: T30552
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1449 <JAF>
A:Cross-references: UNIPROT:O68542; UNIPARC:UPI0000B10PD; EMBL:AF049609; NID:G2935545;
C:Genetics:
A:Gene: gtfN

Query Match 42.6%; Score 3076.5; DB 2; Length 1449;
Best Local Similarity 45.0%; Pred. No. 2.2e-148;
Matches 667; Conservative 205; Mismatches 452; Indels 157; Gaps 35;

QY 1 MEKNLYKLHKVKKQWVAIGTVTLTSLFAGGVVAAADTNNDGTSVQVKNMVPSPDKPD 60
Db 1 MDKKVHKVHKVKKQWTVAVTGLSL-----GAVSAVSLGTNDG-----VQADEHTD 48

QY 61 AQAGQOLAQAMFKAANQ--ADQTATSVSPATDGRVDNQVTPAA-----NQPAAN 109
Db 49 ATVAIPDIITVDTGVSNDITTAQDPTTAAATNDVATD-QATPTATFDLTDTTNTVAAN 107

QY 110 VANQOVANPATDAGALNRQSA-----DTSDDGKAV----- 140
Db 108 -AVDTVATVGTDRAAFTNDTATNDTAVDTTNNNTTDTTNTNRAATERRATGARRGP 166

QY 141 -----POT-----SDPGHLETVGKTYVVDANGORLKN 169
Db 167 TGGRRATPVNGMNNANNTVVVNNDLPATNVNVTGDPSHIKTINGKQYVVEDDGTIRKN 226

QY 170 YSMVIDGKTYFPDGTGEAQT-----LPKTGANODNVPDSQAN-----NQAYSNEA 218
Db 227 VYLERIGGSQYFNAETGELSNOKEYRFDKNGGTSS--ADSTNTVTVNGDKNAFYGTTD 284

QY 219 SSFETVDNYLTADSWYRPKRILKNGSOWASSEGDLRPLMTWPDPAATKAAAYANFWAKE 278
Db 285 KDIELVDGYFTANTWYRPKEILLKDGKEWTASTENDKRPLLTVMWPSKAIQASYLYNMYKEQ 344

QY 279 GI-1SGSVRQNSA--NLDAATQNIQSAIEKKTASEGNTWLRDKMSQFVKSQNSIASIE 335
Db 345 GLGTNTQYTSFSQQTQMDQAALVEQVRIEGRAREGNTDNLRTIKTKNFKVKTQPGWNSTSE 404

QY 336 NETVYPNQDHMOGALLFNSKDTTEHANSWRLNRPNTFTQCKQ--KYF-TTNYAGYEL 392
Db 405 N---LDNNDHLQGGALLYNNDSRTSHANSYRLNRPNTFTQCKNPKYTKDTSNGGPF 461

QY 393 LLANDVNSNFPVQAEOLNHLHYLNMWGDIVMGDKDANFDGVRVDAVDNVNADLLQIQORD 452
Db 462 LLANDIDNSNFAVQAEOLNHLHYIMNIGTITGGSEDEFNFDGVRVDAVDNVNADLLQIASD 521
QY 453 YYKAKYGTQDQNEKAIDHLSILEAWSGNDNDYVKDQNNFSLSDNDQSGMLKAFGYASA 512
Db 522 YFKAKYGADQDQQAIKHLSILEASHNDAYNEDTKGAQLPMDDPMHSLAYSLRPIG 581
QY 513 YRGNLSNIATAGLKNRS--ANPDSDPVPVYFIRAHDSVQTRIAKIIREKLGKTNADGLT 571
Db 582 NRSGEPLISLNSLDRSESGKNSKEMANYAFVRAHDSVQSIIGQIIKNEINPOSTG--N 639
QY 572 NLTLDDLKAPDIYNODMNAIDKUYYPNNLPMAVAMLONKDTTVRVVYGYDMYTONGYM 631
Db 640 TFLDEMCKKAFEIYNKDRMSANKQTYQNIIPSAVALMLTHKDTVPVPRVYGYDMYTDDGQM 699
QY 632 ATKTFFYNAIETLLKGRIKYVAGGQ--AVSY-----KODW--SSGILTSRVYKGSANSASD 683
Db 700 AQKSPYDAIETLLKGRIRYAAAGGDMKVNYIGYNTNGMDAAGVLTSVRYGTGANSASD 759
QY 684 AGNTESTRSGMALLINRPNFRAYRNLTLNMGAAHKSQAYRPLLLSTKDGATYLNDSDV 743
Db 760 TGTAETRNQGMAVIVSNQPALRLTSLNTLNMGAAHNRQAYRPLLLTTNDGVATYLNDSDA 819
QY 744 DSRQVYKTDSONLSPSASELSQVANAOVSGMIQVWVPVGAADNODVRTSPSTQATKQGN 803
Db 820 NG-IVKYTDGNGNLTFSSANEIRGRNPQVDGLVAVVPGVASENODVRVPSKEKNSSL 878
QY 804 IYHOSDALDSQVIEYEGFSNFQAFQSPQYTNVIAKNGDLFKSWGITQFENAPQVVASSE 863
Db 879 VYESNAALDSQVIEYEGFSNFQDPVNPQSYTNKTAENANLFXSWGITSFEPAPQVSSD 938
QY 864 DGTFLDSVILNGYAFSDRYDLAMSNNKYKSGQDLANAIKGLQSAIGIKVLSDLVNPQLYN 923
Db 939 DGSFLDSVIQNGYAFSDRYDYGMSKDKYKSLADLKAALKSLHVGISAIADWDPQIYN 998

QY 924 LPGAKEVTVATRVNQYQAKSGATINKTPYVANTRSYG--DYQEOYCGKELDLQKLYPLRF 982
Db 999 LPGDEVVTVATRVNNYGETKDGAIDHLSYAAKTRTFGNDYQGYGGAFLDELRLYPOIF 1058
QY 983 STKQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKYVYLSEB--NKYLLNADGKLFPLTVL 1041
Db 1059 DRVQISTGKRWTTDEKITKNSAKYNNGTNLLDRGSEYVLKNGLNGYGTNGGKSLPKV 1118
QY 1042 NNTYQGPQVSANG-----FISKNGGIHVDKXGQEVKNRP--KEISGSWYVYFSDGSMATG 1095
Db 1119 GSNQSTNGDNQNGDGSKFEKRLFSVRYRNNQYAKNAFIKDNNDGNVYVYFDSNGRMVAG 1178
QY 1096 KTKIGNDTYLPWPNKGKOLKEGVWYDGK--KAYYYDDNGRTWTN-----KGFB 1141
Db 1179 EKTIDGKQYFFLANGVLEDDGYRQNRGQVFFYDQGVNLNANGKODPKPDNNNNASGRNQ 1238
QY 1142 FRVDGQDKWRYFNGDGTIAIGLVSLDNRITLYFDAYGVYQVKGQTVTINGKSYTFDADQDGL 1201
Db 1239 FVQIGNVWAYVDGNGKRVTHQNINGQBELFFDNNNGVQVKGRTVNGGAIRYYDANSSEM 1298
QY 1202 VOTDANPAQOAGKWLGDNQWGY-RKDGQLLTGEGTIDGQVFFQDNGVGVKGGTAT 1260
Db 1299 ARNRFAETEP-----GVWAYFNNDGTAVKGSQNINGQDLYFDQNGRQVKGALA- 1346
QY 1261 DASGVLYRFDYDRDQGHQVKGWYSTSDNNWVYVNESGOVLTLGTLTIDGQTVYFDDKGIQAK 1320
Db 1347 NVDGNLRYDVNSG-ELYRNRPHEDGSMYFDGNGVAKGMVNINGQNLFFDNNKGQIK 1405
QY 1321 GKAVWDEGNLRYFDADSGNMLDRWKNVD-GNYYFNRNG 1360
Db 1406 GHLV-RVNGVRYFDPNSGEMAVNRWVEVSPGWVYFDGEG 1445

RESULT 5
T30858
glucosyltransferase - Streptococcus salivarius

Db	774	GGQDMKVTKLNGYBMSVRYCKGAEEANQLGTAETRNQGMVLVTANRPDMKLGANDRLV	833
Qy	712	LNMGAHKSQAYRPLLSTKDCIATYLNDSVDVSROYKYTDSQGNLSFSAELOSQVANAQ	771
Db	834	VNMGAAHKNQAYRPLLSTKGLATYLNKSDVPAGLVRYTDNQNLTFPTADDIAGHSVTE	893
Qy	772	VSGMIQWVPVGAADNQDVRTPSPQATKDG-NIYHQSDALDSQVIYGFNFQAFAPASQ	830
Db	894	VSGYLAVWVPVGAENQDQARTKAS--STPKGEQVPFESSAALDSQVIYGFNFQDFVKTP	951
Qy	831	DOYTNVIAKNGDILFKSWGITQFEMAPQYVSSEGGTFLDLSVILNGYAFSDRYDLAMSKNN	890
Db	952	SQYTRNVIAQNAKLFKSWGITSFFEPAPQYVSSQGTFLDLSIIENGYAFEDRYDIAMSKNN	1011

[illegible]

QY 951 PYVANTRSYG-DYQEQYGGKFLBDBLQKLYPRLFSTRQISTGKPIDPSVKITNWSAKYFNG 1001

DB 1072 LYAAKTRTFGNDFFQKYGGAFLDELKAKYPAIFERVQISNGRKLTTNEKITQWSAKYFNG 1131

QY 1010 SNILGRGAKYVLSEG--NRYLNLAADGKLFPLVNLNTYQPPQVSANGFISKNGGIHYLDK 1067

DB 1132 SNIQGTGARVWLQDNATNQYPSVRAGOTFLPKQMT-----EITSGGFRRVGDDVQYLSI 1185

QY 1068 NGQEVKNRFKEI-SGSWYFYFSDGKMATGKTKIGNDTYLFMPNGKQLKEGV----- 11111

DB 1186 GGYLAKNTFIQVGANQWYYFDKNGNMVVTGEQVIDGKKYFFLDNGLQLRHLVLRQGS DGHVY 1245

QY 1118 WYD--GKRAY-----YDDNGRTW-----TNKGFVE 1141

DB 1246 YYDPKGVQAENGFYDFAGPRQDVRYFDGNGQMYRGLHDMYGTTFYFDEKTIQAQDKFIR 1305

QY 1142 FRVDGQDK-----WRYFNGDGTIAIGLVSLDNRTLYFDAYGYQ 1175

DB 1306 F-ADGRTRYFIPDTGNLAVNRFAQNPENKAWYLLDSNGYAVTGLQTINGKQYFFDNEGRO 1364

QY 1180 VRGQTVTINGKSYTFDADQGBLVQ-----TDNANPAFQGGAGWKLGLGDNQWGY-RKRDGQLL 1234

DB 1365 VRGHFVTINNQR YFLDGDSGEIAPSRFVTEN-----NKWYYVDGNGKLV 1408

QY 1235 TGEQTIDGQKRVFFQDNGVQVKGGTATBASGVLFYDRDQGHQVGKGWYSTSDDNWVWNE 1294

DB 1409 KGAQVINGNHYYFNNDYSQVKGAWANG-----RYYDGDGSGQAVSNQFIQIAANQWAYLNQ 1463

QY 1295 SGQVLTGLQTFIDGQTVYFDDKGIQAKGKAVWDENGNLRFYFDADSGNMLRDRW-KNVDGNW 1355

Db 1464 DGHKVTGLQINNNKVYFSGNGAQVKGKLL-TVQGKKCYFDAHTGEQVNNRFVEAARGCW 1522

QY 1354 YYFNRLGA 1362

DB 1523 YYFNSAGQA 1531

RESULT 6

A381/5
glucosyltransferase precursor - Streptococcus sobrinus

C;Species: Streptococcus sobrinus
C;Date: 28-Aug-1992 #sequence revision 28-Aug-1992 #text change 15-Mar-2004

C;Accession: A38175
R;Abo, H.; Mtsumura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.

A: Title: Peptide sequences for sucrose splitting and glucan binding within Streptococcus

A;Reference number: A38175; MUID:91123227; PMID:1704006
A;Accession: A38175

A;Status: preliminary
A;Molecule type: DNA

A;Residues: I-1592 <ABO>
A;Cross-references: UNIPARC:UPI0000012BCB2; GB:D90213; NID:q217032; PIDN:BAAL424

F:1093-1112/Domain: cpl repeat homology <CP1>
F:1222-1241/Domain: cpl repeat homology <CP2>

F;128/-1306/Domain: cpl repeat homology <CP3>

F:1330-1351/Domain: cpl repeat homology <CP4>
F:1352-1371/Domain: cpl repeat homology <CP5>
F:1402-1420/Domain: cpl repeat homology <CP6>
F:1465-1484/Domain: cpl repeat homology <CP7>
F:1513-1532/Domain: cpl repeat homology <CP8>

Query Match 41.8%; Score 3023; DB 2; Length 1592;
Best Local Similarity 41.3%; Pred. No. 1.4e-145;
Matches 669; Conservative 221; Mismatches 436; Indels 292; Gaps 36;

QY 1 MEKNRYKLHKYKQKQWVAIGVTVTV-LSPLAGQGVVAADTNNDGTSVQVKNKVPSPKPF 59
DB 1 MEKNVRFKWHKYKKEKVVTLVSVASATMLASALGASVASADTDTASDSSNQ-T-VVTGD--- 55

QY 60 DAQAQNGQLAQAAMFAAQAQDQATQTSQVSPATDGRVDMQVTPAANQPAANQDV---- 115
DB 56 --QTTNNQATDQTSIAA-----TATSEQSASTDAATDQ--ASAAPQTCQTASTDTAAQT 106

QY 116 ---ANPA-----TDAGALNRQSAADTSTDGKAVPQTSQDPGHLETVGCKTYV 160
DB 107 TTNAEAKWVPTENENQGFDEMLBAKNVATAESD--SIPSDLAKSNVKKQVDKY 164

QY 161 DANGQRLKNYSNVIDKTYVFGQGTG-----EAQTDLPKGTQANQDNPDPSVQANNOAYS 216
DB 165 DQGNVKNFVAVSGDKIYFD-ETGAYKDTSKVDADKSSAVSQNA-TIPAANREAYST 222

QY 217 EASSFETVDNYLTADSWYRPRKILKNGOSWQASSEGDLRPLMTWPDAAATKAAVAFWA 276
DB 223 SAKNFEAVDNYLTADSWYRPRKILKNGKWTWESGKDDFRPLLMAWVPTETKRNVTYNN 282

QY 277 KEGLSGSY--RQNSANLDAATQNTQSALEKKIASEGTNWLRDKMSQFVKSONQWSTAS 334
DB 283 KVVGDKTYTAETSQADLTAAAEVLQVAREQKITSENNTKLREAISAFVKTQPMNGES 342

QY 335 ENETVVPNDHMOGQALLFPNSKD--TEHANSWRLNRRNPTFQTGK-QKYFTTN----YA 388
DB 343 EK----PYDDHLQNGALLFDNQTDLTPTQSNYRLNLRPTNTQTSLSRFTYFNPDP 398

QY 389 GYELLANDVNSNPVQAEQLNHLHYLMNWGDIYVNGDKOANFDGVRVDVADNVNADLLQ 448
DB 399 GYDFELLANDVNSNPVQAEQLNHLHYLMNWGDIYVANDADANFDSIRVDAEDNVADQLQ 458

QY 449 IQRDYKAKYGTQDQEKNAIDHLSILEAWSGNDYVYKQDNFSLSDNDQBSGMLKAFG 508
DB 459 ISSDYLKAAAGYDKNNKANNHVSIVEAWSNDTPLYHDDGNLMMNDKFLSLMWSLA 518

QY 509 YASAYRGNLSNLATAGLKNRSANP--DSDPVNPVFIKRAHSDSEVQTRIAKIIREKLK 567
DB 519 KFTDVRSGNLPLHNSLVDREVDREVEVTPVPSYFARAHDSEVDIIRDLIAEI-NPNS 577

QY 568 DGLTNLTLDLNLKAFDIYNQDMNATDKVYYPNNLPMAYMWLONKDTVTRVYVYGDYTDN 627
DB 578 FGYSFTQBEIDQAFKIYNEEDLKQDKKYTHYNVPLSYTLTLTKNGSIPRVYVYGMFTDD 636

QY 628 GQYMATKTPFYNAETILKGRKYVAGGQAVSYKQDWSSGILTSVRYGKGANSADAGNT 687
DB 637 GQYMAKNTVYDAIESLLEKARMKYVAGGQAWQNYQINGEILTSVRYGKGLKQDKGDA 696

QY 688 ETRNSGALLINRPNFRA-YRNLTLNMGAAHKSQAYRPLLLSTKDGATYLNDSDDVD-S 745
DB 697 TTRTSGVGVVMGNQPNFSLDGKVALNMGAAHANQEYRALMVSTKDGAVATYATDASKA 756

QY 746 ROYKYTDSQGNLSFSAEQLSVANAQVSGMIQVWPVGAADNQDVRTSPSTQATKGNITY 805
DB 757 GLVKRTDENGYLTYFLNDDLKGVANPQVSGFLQVWPVGAADQDQIRVAASDTASTDGKSL 816

QY 806 HQSDALDSQVIYEGFSNFOAQAQSPQDYNNAVIAKNGDLFKSWGITQPFEMAPQVYSSDG 865
DB 817 HQDAAMDSRWMEGFSNFOFATKEEYTNVVIANNVDFKVSFSGITDFEMAPQVYSSD 876

QY 866 TFLDSVILNGYAFSDRYDLAMSKNNKYSQDLANAIKGLQSAGIKVLSLDVLPNQLYNLP 925
DB 877 QFQDSVQNGYAFTRYDYLGMSEKANKYGTADQLVKAIKALHAKGLKVMADWVPDQMYTFP 936

QY 926 GKEVVTATRVNOYQAKSGATINKTPYVANTRSYG-DYOEQYGGKFLDLOKLYPRLPST 984
DB 937 KQEVVTITRTDFKPKITAGSQINHSLSYVDTFKSSGDDYQAKYGGAFDLDELKCKYPELFTK 996

QY 985 KQISTGKPIDPSKIVITNWSAKYFNGSNILGRGAKYVLSE--GNKYLNILADGKLFPLTVLN 1042
DB 997 KQMTSGQALDPSVKIKQWSAKYFNGSNILGRGADYVLSQVSNKYFNVAVDTLFLPSSLL 1056

QY 1043 NTYQOQVSNANGFTSKNGGIHYLDKNGQEVKNRFBKISGSWYFSDSGKMATGKTGKIGND 1102
DB 1057 GKVVESGIRYDQ---KGYIYNSSATGDQVKASFTTEAGNLYYFGKDGVMVTGAOTINGA 1112

QY 1103 TYLFWPCKQLKEGVVWDGK-KAYYYDDNGRTWNK-CFVEFRVDGQKWKWYENGSGTIA 1160
DB 1113 NYFFLENGTALRNTIYTDAQGNSHYYANDGKRYENENGYQQF---GND-WRYFK-DGNWA 1167

QY 1161 IGLVSLD----- 1167
DB 1168 VGLTVDGNVQYFDKDGVOAKDKIIVTDPDGKVRYPDQHNNAVNTFTIADKTHWYILGK 1227

QY 1168 -----NRTLFPDAYGVYKQGTVTIN-GKSYTFDADQDGL----- 1201
DB 1228 DGVAVTGAQTVGKQKLYPEANGQOVKGFVTSDEGKLYFYDVSQDMWMTDTFIEDKAGNW 1287

QY 1202 ----- 1201
DB 1288 FYLGKQGAOAVTGAQTIHQKLYFKANGQOVKGDIVKPGDKIRIYDAKSGQVFNKTVKA 1347

QY 1202 -----VOTDNANPAP-----OCQ-----AGMKLLGDNQWGYRKD 1230
DB 1348 ADGTYVIGNDGVAVDPSVVKVGQTFKDGASGALRPNLKGQLVTSQGWYETANHDWYIQS 1407

QY 1231 GQLLTGTEQTDQGVFFQDNGVQVKGQGTATDASGVLFYDRDQGHQ----- 1276
DB 1408 GKALTGTEQTINGQHLFYFKKGQHVKGQVLVTGTDGKRVYDANSQDAFNKSVTVNGKTY 1467

QY 1277 -----VGKQWYSTSDNNWYVYNESGOVLACL 1302
DB 1468 FGNDDTAQTAGNPKQGTQFDGSDIRFYSMEQQLVTGSGWYNAQOWLYV-KNGKVLTLG 1526

QY 1303 QTIDQGTVYFDDKGIQAKGKAVMDENGRLRYPFDADSGNMLDRWKNVKNVGNWYFNRNG 1360
DB 1527 QTVGSQVYFDENGIQAKGKAVRISDGKIRYFDENSGSMITNQWKEVNGRYYYFGNDG 1584

RESULT 7
B31135
gfB protein precursor - Streptococcus mutans
C:Species: Streptococcus mutans
C:Date: 23-Oct-1990 #sequence_revision 23-Oct-1990 #text_change 09-Jul-2004
C:Accession: B31135; A33128
J:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
J: Bacteriol. 169, 4263-4270, 1987
A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.
A:Reference number: A33135; MUID:87308013; PMID:3040685
A:Accession: B31135
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1475 <SHI>
A:Cross-references: UNIPROT:P08987; UNIPARC:UPI000014D972; GB:M17361; NID:g153639; PIDN
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
A:Experimental source: strain GS-5
A:Reference number: A33128
A:Accession: A33128
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-171,173-641,'N',643-1475 <SH2>
A:Cross-references: UNIPARC:UPI000017AC5E
A:Experimental source: strain GS-5
F:1096-1115/Domain: cpl repeat homology <CP1>
F:1224-1243/Domain: cpl repeat homology <CP2>
F:1289-1308/Domain: cpl repeat homology <CP3>


```
QY 675 GKGANSADAGNTETRNHSGMALLNNRPN--FRAIRNLTLNMGAAHKSQAYRPLLLSTKD 732
DB 676 GKGANSADAGNTETRNHSGMALLNNRPN--FRAIRNLTLNMGAAHKSQAYRPLLLSTKD 732
QY 733 GIATYLNDSVDSDROYKYTDSQGNLSFASSELQSVANAQVSMIQVWVVPVGAADNODVRT 792
DB 827 NLSYV----DNDNAPVKYTNDDQGLLIIDNTEIYGVNRPQVSGFLAVWVVPVGAADNODVRT 882
QY 793 SPSTQATKQGNIIHQSDALDSQVIYEGFSNFQAFAPQSDPOYTNAVIAKNGDLFKSWGITQ 852
DB 883 LSDDTAHDGKTFHNSAALDSQVIYEGFSNFQAFATNTEDYTNVIAKNGDLFKSWGITQ 942
QY 853 FENAPQVYSSDGTFLDSVLNGVAFSDRDYDLAMSKNKGKSGQDLANAIKGLQSGAGIKV 912
DB 943 FQAPQVRSSTDSFLDSIIQNGYAFTRDYLGTGTPTKVTGVDQLDADAIKALHANGIOA 1002
QY 913 LSDLPVNLNLPQKGVVTRVNOVGOAKSGATINKTPVANTRSVGDYQEOYGGKFLD 972
DB 1003 IADWVDDQIYNLPQELATVTRNSYGDQDNTSDIDQSLVIOISRGGKQYQATYGGAFLS 1062
QY 973 DLQKLYPRLPSTKQISTGKPIDSPSVKITNWSAKYFNGSNILGRGAKYVLSEGN-----KY 1027
DB 1063 DIQKYPALFETKQISTGLPMDPSQKITWSGKYFNGSNIOGKAGYVLKDSGTDQYKYV 1122
QY 1028 LNLADGKFLPTVLNNTYQGVQVANGFISKNGGIIHYLDKNGQEVKNRP-KEISGWSYVF 1086
DB 1123 TSNNNRDLFLPKQLTD-----DLSETGFVRDNGMVYVYTLSGYLARNTFIQDDNGNYVF 1177
QY 1087 DSPGKWATGKTKGNDTYLFWPKNQKQKGVWY--DGKAYYYDDNGRTWTKGVEFRV 1144
DB 1178 DSTGHLVTGQNTNNHHYFLPLNGIELVQSLQVADG-STIYFDQGR-----QVFNQYIT 1232
QY 1145 DGQDKWRYFNGDGT-IAIGLVSLDNRTLYPDAYGYOVKGQTVT-INGKSYTFDADQGLV 1202
DB 1233 DQGTAYYFQNDGTWYTSFGTEIDGHQYFYKNGTVQVGFVSDTDGHVYLEAGNGN-V 1291
QY 1203 QTNANPAPQOGAKWLLGNQNGYRKDGQGLLTGEGOTIDGQVFFPDNGVQVKG-----1257
DB 1292 ATORFAQNSQGO--WFYLG-----DGIALTGLQTINGVQVYFVADGHQSGKDFITIQ 1342
QY 1258 -----TATDASGVLYRFDYR-----DQGHQVKGWSTSDNNVYVNESQVLTGLQ 1303
DB 1343 NHVLYTNPLTGATTCMQQIQGDKIFVFDNTGNMLTNOYVOTLQDQWHLSTQGPADTGLV 1402
QY 1304 TIDGQTVYFDDKGIQAKGKAVMDENGNYRF-DADSGNMLRDRKKNVGDNGWYVFNFRN 1359
DB 1403 NINGNLKYFQANGRVKQGVPTDPTNVSYMMATDGSVAFFNDYFYQGGWYLTDSN 1459

RESULT 12
JCS473
dextranucrase (EC 2.4.1.5) - Leuconostoc mesenteroides
C:Species: Leuconostoc mesenteroides
C:Date: 07-Jul-1997 #sequence_revision 29-Aug-1997 #text_change 09-Jul-2004
C:Accession: JCS473
R:Monchois, V.; Willemot, R.M.; Remaud-Simeon, M.; Croux, C.; Monsan, P.
Gene 182, 23-32, 1996
A:Title: Cloning and sequencing of a gene coding for a novel dextranucrase from Leuconoc
A:Reference number: JCS473; MUID:97136686; PMID:8982063
A:Accession: JCS473
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1290 <MON>
A:Cross-references: UNIPROT:Q48756; UNIPARC:UPI000017ABDF; GB:U38181
C:Comment: This enzyme catalyzes the transfer of D-glucopyranosyl units from sucrose and
C:Genetics:
A:Gene: dera
C:Keywords: glycosyltransferase; hexosyltransferase
F:78-970/Domain: catalytic #status predicted <CAT>
F:922-1290/Domain: glucan-binding #status predicted <GCB>

Query Match 35.9%; Score 2593.5; DB 2; Length 1290;
Best Local similarity 42.2%; Pred. No. 6.8e-124;
```

```
Matches 543; Conservative 228; Mismatches 395; Indels 121; Gaps 24;
QY 149 HLETVDGKTYVVDANGORLKNYSWIDGKTYYPFDGOTGEAQDTLPTKQGANQVNPDSYQ 208
DB 3 NVKHVDGKTYFYDGDQPKNFYTIIDGKPYFPDKDTGALSNN-----DKQVVELFS 55
QY 209 ANNQ-----AYSNEASSFETVDNYLTADSWYRPRKILKNQSQWASSEGDLAPILMTWMPDA 265
DB 56 IGKHNNAVNTSNDNFQLEHGLTASSWYRPRKILKNQSWAPSTVTDPRLLMAWMPDK 115
QY 266 ATCAAAYANFWAKGLISGSY-----RONSANLDAATQNIQSAIEKKIASEGNTWLRDMS 321
DB 116 STQVTVLNTYMKQGLLSGTHFSDNENMRTLTAAMQAOVNIIEKKIQLGNTDMLTKAMT 175
QY 322 QFVKSQNSIASENETVYPNODHMOGALLFNSKDTTEHANSDWLRNRPFTQTK-- 379
DB 176 QYIDAQPNWIDSEAK-----GDDHLOGGALLYTNSDMSPKANSDYRKLSTPKNKGQIA 231
QY 380 QKYFTTNYAGYELLANDVDNSNPVVOAEQLNHLHYLMNMGDIWMGDKDANFDGVRVDAV 439
DB 232 DKY---KOGGFELLANDVDNSNPVVOAEQLNHLHYLMNMGDIWMGDKDANFDGVRVDAV 288
QY 440 DNVNADLLQIQDYKYAKYGTTONKVAIDHLSILSAWGSNDNDYVKDQ-----489
DB 289 DNVNADLLQIAGEYAKAAYGVD---KMTREINIIYQFWKTGE---MKIQTMSKHMATSKL 342
QY 490 -----NFSLSIDNDORSMLKAFGVASAYRGNLSNLATAGIKNSANPDSDPVNP 539
DB 343 SMDFFPLHLAIKYALNPNPNDRSGLEPTREHSL-----VKRITDDKENVAQPN 389
QY 540 YVFIRAHDSVQTRIAKIIREKLGKTNADGLNLTLDLNLKAFDIYNQDMNATDKVYYPN 599
DB 390 YSFIRAHDSVQTIADIIKINPASTGLSDSTVTLQKQAFDIYNADDELKADKYVTPY 449
QY 600 NLPMAAYAMWLNQKDTVRYVYGDYMDYNDQYMATKTPFNATETLKGRIKYVAGQA---657
DB 450 NIPASYALLLPNKDTIPRIVYGDYMDTDCQYMAKOSPPYQAIADALKARIKYAAGQTMK 509
QY 658 VSYKODWSSGILTSVRYGKGNASDAGNTETRNHSGMALLNNRPNFRA--YRNLTLNMG 715
DB 510 MYFFPDEQ--VMTSVRYGKGNATASDGNQETRYQGILVNVNRPDLKLSDDKEDVMDMG 568
QY 716 AAHKSQAYRPLLLSTKQGIATYLNDSVDSDROYKYTDSQGNLSFASSELQSVANAQVSM 775
DB 569 AAHKNQDYRPLVLTTSKGLKYVTDANAPVVR---TDANGQLTFKADMYGVNDPQVSGY 625
QY 776 IQWVPVGAADNODVRTSPSTQATKQGNIIHQSDALDSQVIYEGFSNFQAFAPQSDPOYTN 835
DB 626 IAAWVPVGAENQDARTKSETTQSTDGVSYYHNSAALDSQVIYEGFSNFQDFTPTTDEFTN 685
QY 836 AVIAKNGDLFKSWGITQPEMAPQYVSSDGTFLDSVLNGVAFSDRDYDLAMSKNKNKYGSK 895
DB 686 IKIAQVNLVFKDNGITSEFMAPQYRASSDKSLDAIVQNGYAFTRDYLIDGYNPTPKYIGTA 745
QY 896 ODLANAIKGLQSGAGIKVLSDLVPLNQLYNLPKGVEVVTATRVNQYQAKSGATINKTPYVAN 955
DB 746 DNLLDALRALHQQGIIQAINDMVDPQIYNLPDSQLVTAIRTDGSGDHTYGSVIDHTLYASK 805
QY 956 TRSYGDYQEOYGGKFLDQLKLYPLKSTQKQISTGKPIDSPSVKITNWSAKYFNGSNILGR 1015
DB 805 TVAGGIYQQYGGAFLEQLKTOYPLQFOQKQISTQPMNPDIQIKSWEAKYFNGSNIQGR 865
QY 1016 GAKYVYLSE--GNKYLNLADGKFLPTVLNNTYQGVQVANGFISKNGGIIHYLDKNGQEVK 1073
DB 866 GAWYVVKDNGTQQYFNVSDAOTFLPKQLLG-----EKAKTGFVTRGKETSFYSTSGYQAK 920
QY 1074 NRFKEISGWSYFSDSGKWATGKTKIGNDTYLFPMPKNQKKEGVWVDGKAYYYDDNGRT 1133
DB 921 SAFICDNGWYVYFDDKGVNMGVNGVINGVINYFLPNGIELQDAYLVHDGMYYNNIGKQ 980
QY 1134 WTKNGFVFRVDGQDKWRYFNGDGTIAIGLVSLDNR-----TLYFDAYGYQVKGQTVT-I 1187
DB 981 LHN-----TVYQDKQKNFHYFFEDGHMAQGIIVTIIQSDGTFVTVQYFDENGKQKGVAVKGS 1036
```

QY	1188	NGKSYTFDADQGDLV-----QTDN-----ANPAPQGQAG	1216
DB	1037	DGHLHYFDGASGNLPSKWSRLADGSLWYDEKGNNAVTKGTINNQTVYFNDGGRQIKNN	1096
QY	1217	WKLGLDGNQGY-RKDQLLTGSGTIDGQKVFQDNGVQKGGTATDASGVLFYDRDOGH	1275
DB	1097	FRELADGSLWYLNKKGAVTGEQINGQTLFYFGDGRQPKGTHINATGESRYYPDPSGN	1156
QY	1276	QVKGKMYSTDNDWVYVNESGOVLTLGLQIDGQTVYFDDKGIQAKGKAVMDENGLRYPD	1335
DB	1157	MITDRERVDNQWAFYGDYVAVTGDRIIKGNLYFNQNGIQMKGH-LRLENGIMRYVD	1215
QY	1336	ADSGNMLRDRWKNV-DGNWYFYFNRNGL	1361
DB	1216	ADTGELVRNRFVLLSDGWSVYFQDGV	1242
RESULT 13			
B97033			
uncharacterized protein, related to enterotoxins of other Clostridiales [imported] - Cl			
C:Species: Clostridium acetobutylicum			
C:Date: 14-Sep-2001 #sequence_revision			
C:Accession: B97033			
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,			
J. Bacteriol. 183, 4823-4838, 2001			
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cl			
A:Reference number: A96900; MUID:21359325; PMID:21359325			
A:Accession: B97033			
A:Status: preliminary			
A:Molecule type: DNA			
A:Residues: 1-2817 <KUR>			
A:Cross-references: UNIPROT:Q97K42; UNIPARC:UPI00000CA0A0; GB:AE001437; PIDN:AAK79053.1;			
A:Experimental source: Clostridium acetobutylicum ATCC824			
C:Genetics:			
A:Gene: CAC1079			
Query Match 7.5%; Score 543; DB 2; Length 2817;			
Best Local Similarity 22.7%; Pred. No. 4.2e-19;			
Matches 357; Conservative 190; Mismatches 542; Indels 486; Gaps 89;			
QY	134	STDGKAVPQSDOPGHELVTDGKTYVVDAGORLKNYSWVIDGKTYFF-----	181
DB	451	ANDGKLLT-----GLQVINGSYFDTNGIRLVSRLWITIDGKDYFNQDGLTNDWIN	503
QY	182	DG-----QTGEAQT-----LPKTQANQDNVP-----DSY	207
DB	504	YDGKYFYISGVKQTLGNIDGNYFYFSDSGIMQTLGKIDGKTYFPGDNGIRQIGWITY	563
QY	208	QANNQAYSNEAGSFET---VDNYLTADSWYRPR-----KILK-----NQGSWQASSE	251
DB	564	Q-NKKYFNSDGSMTDLKIYSYSTSPYNVHYQYGFNDGKLLTLGLTIKNTYFYFDSN	622
QY	252	GDLRPLMTWPDAAKAYAFWAKEGLISGYSRONSANLDAATONIOSAEKKIASE-	310
DB	623	G-----ISQMGWNIDKDFYN---SNSIMTENWVINDKYYFINNVKQTCGFQYINGKY	675
QY	311	-----GNTNMLRD---KMSQFVKSQ-----NOMSIASENE	337
DB	676	YFDPDPMQGTGQTISGNTYLLDDNGVKTGQVHTIKGDYFPGNGVMINTW-VFDNDK	734
QY	338	TVYPNODHMQGGALLFSNS-----KOTEHANSWRLLN-----RNPTFQT-----GKQ	380
DB	735	TVYYINGNNQMTGAISINNHYFGDDNGIMQGTQWRINGRITYFDDNNGAAKTGLVTEGKT	794
QY	381	KYFTTNYAGYE---LLAND---VDNSNPVQAEQLNH---LHLYLNMWGDIVMGDK---DA	429
DB	795	YYPNTYAYLDTGFIYFNNNYFPLDN-NGVVRTGWINYSNNRYLSDSTGVRVTGFTIDG	853
QY	430	N---PD--GVRVDAVDNVNADLLQIQRDYKAKYGTQDQEKNAIDHLSILEAW-----SGN	480
DB	854	NKXYFSDSSGAMCTSFITVNGN-----TYGF---SKDGI-----MLTGQHTILSSN	895

QY	481	DNDYVKQDNFSLSDINDORS-----MLKAFGYASA---YRGNLSNL	520
DB	896	YSSY-----NIYFNSDGAQKGFYTLGKTYFEPNYGYMLLGYNYINGKYYFNDGVI	951
QY	521	ATAGLKNRANPDSDPVENVV--PIRAHDSVQTRIAKIIREKLKTNAD-----	568
DB	952	QTGWTDSSKXYLDPGAAVTGFTQINGDKYFNSSGIMQTLVYVNPDPYGFDDNGHI	1011
QY	569	--GLTNLT-----LDDLKA-----FDIYNQDMNATDKVY--PNNLPMAY	605
DB	1012	LTGHSINGIYIYFYSTGKAKQKGFVYTLGKTYFNTNMYTGFVNANNLLYFDPNEGVMQT	1071
QY	606	AWMLQNKDVTTRVY-----GMYTDNGOY-MATKTPFNATETLLKGRIKYVAGQ	656
DB	1072	GMINTNSN---RYFYSATGASVTGFTIDGNKYCFDSNGAIYDVVTI-----NGS	1119
QY	657	AVSYKQDNSSGIL---TSVRYKGKANSASDAGNTETRNS-GMALLNNRPNFRAYRLT	711
DB	1120	TVGFNTD---GIMLTGMQTIIRNRYSSYF---NTYFNSDGT-----KTGFFYLNKT	1168
QY	712	LNMGAAHKSQAYRPLLSTKDGIAIYLNDSVDVSRQ-YKYTDSQGNLSPSASELOSANA	770
DB	1169	Y-----YFNPSDGRMLQGYQYIN--GNHYVFAD-----	1195
QY	771	QVSGMIQV-WV-----PVGADNDQVNTSPSTQATKD-----GNIVH	806
DB	1196	---GTMQGTWITNGSSKTYLDPGSAVT-GLQTINGNKYCFDSNGILQHNGIFYIGNTY	1251
QY	807	QSDALDSQVIYEGFSNFQAF-----QSPQYTNNAVAKNGD--LPKSWGITQFEMAPQ--Y	859
DB	1252	GSD--NNGIMLTGLQINGLYCFNSDGSVKTLGTYLKYTFYFDSYSVSGGQNNNTY	1309
QY	860	VSSDGTFLDSVILNGYAFSDRYDLAMS-----KNKK-----YGSKODLANA---	901
DB	1310	YFGNDGTMTQGWVNYGY--RYLNDSGIKVYTGWTIDGNKYFYDYGYAKTGIVNIDGN	1366
QY	902	IKGLOSAGIKULS-DLVPNQLNLPKKEVVVTRV-----NOYGOAKSGA-TINKT	950
DB	1367	YGFNNSGVMLTGMWHINGSTYFNSNGIANTGTYLTKTYFDFSYGRMQISMTINGT	1426
QY	951	PV-----ANTRSYGDYQEOYGGKFLDLOKLYPRLPSTKQISTGKPIDPSV	997
DB	1427	SYFYFANGVMKTSSTDSPNTLAVGWVDSY-----YYQYLLNAAAGTK--LTG-----L	1471
QY	998	KITNWSAKYFNGSNILGRGAKYVLSEGNKYLNADGKLF--LPTVLNNTY--GQPOVSAN	1053
DB	1472	QTIDGNTYFDSNGIMQGT--IITINGNRGYGFGVNGVMLYGLQFINNNTYYSNGYSQT	1529
QY	1054	GFISKNGGIHYLDKNGQEVKRFKEISGSWYFSDGKMATGKTIGNDTYL--PMPNGK	1111
DB	1530	GFVTLSGNTYFDSYG-EMRIGLTYINNYYFNSKGMETGWI-----SYLRYANPNGI	1583
QY	1112	QLKEGVWVDGKAYYYDDNNGRTWTWKGFVEPRVDGDKWRYFNGDGTIAIGLVSLDNRTL	1171
DB	1584	LUTGQTING-KTYFNSDGSLLYDQYINGSYYGDK-----NGWMLYGLQITCGNTY	1636
QY	1172	YFDAYGYQVKGQTVTINKSYTFDADQGDVLVQTDNANP-----APQG--QAGMKLLGDNQ	1224
DB	1637	YLSNGISQSG-FITLNGKTYFYFDSYGMRTQINNNNNYFPGDNGTLTQGWISQDNLR	1695
QY	1225	WGYRKDGQLLTGEQITIDGQKVPF-----ODNGVQVKGGT--ATDASGVL-----RF	1268
DB	1696	YYANSSGVCLTGLQITIDGKYYFNARMETGLVINNTYFDFDNDGTLLYSWHNINGRM	1755
QY	1269	Y-----DRDQG-----HOVGKWTSTSD	1287
DB	1756	YCFNTDGTVKTGWINYLGRSCYLDSSQGLSTGLLTIGHNIYYFGSDYKMTGW-TSSG	1814
QY	1288	NWVYVNESQVLTGLQITIDGQTVYFDDKGIQAKGKAVMDENGLRYFDDAGSNMLRDRWK	1347
DB	1815	SKYFYNESGIMLTGFTIDGNTYFYFDSYGNSTTGTG--SINGNCYGFN-DDGIMLTG-WQ	1870

Db 2035 GFEYFAPANTYNNIEGQAIYVQSKFLTLNGKKYFDDNNSKAVTGWQITD-----SK 2086
QY 1005 KYFNGSNILGRGAKYVLSGNYK-----LNLA-----DGKFL-----PTVL 1041
Db 2087 KYFYNTNTAAEATGWTIDGKKYFNTNTAAEATGWTIDGKKYFNTNTAIASTGYTII 2146
QY 1042 N-----NTYGOPOVSANGFISKNGIHYLDKN-----GOEV--KNEFKKISGSWYF 1086
Db 2147 NGKHFFYNTDGMQIGV--FKGPGFYEYFAPANTDANNIEGQAIYQNEFLTLNGKKYF 2204
QY 1087 DSDGKMATKTKIGNDTYLFPMPGKQKBEVWYDGKAYYDDNGRTWTKGFEVFRVDG 1146
Db 2205 GSDSKAVTGWRLNNKKYFENNA-----IAIHLCTIN 2239
QY 1147 QDKWRYFNGDGTIAIGLSLDNRTLYFDA-----YGYQVK 1181
Db 2240 NDKY-YFSYDGLQNGYITERNPFYFDANNESKMVTGVFKGPGFYEYFAPANTENNIE 2298
QY 1182 GQTV-----TINGKSYTFDADQGLV--QT-----DNANPAQOAGWKLGDQW 1225
Db 2299 GQAIYVQNKFLTLNGKKYFDDNDSKAVTGWQITIDGKKYFNLTAEATGMQITDGKKY 2357
QY 1226 GYR-KDQGLLTGEQITIDGKVPFQDNGVQVKGGTATDASGVLRFYDRDQGHQV----- 1278
Db 2358 YFNLTAEATGWTIDGKKYFNTNTFIASGT-YTSINGKHFFYNTDGMQIGYFKGPN 2416
QY 1279 -----KGMYSDDNNVYNESGGVLTGLQITIDGQTVYFD-DK 1315
Db 2417 GFEYFAPANTDANNIEGQAIYQNKFLTLNGKKYFGSDSKAVTGLRTIDGKKYFNTNT 2476
QY 1316 GIQAKGKAVDE-NGNLRFPDADSGNMLDRKNVGNWYFNRLG 1361
Db 2477 AVAVTG---WQTINGKKYFNTNT-SIASTGYTIISGRKHFFYNTDGI 2519

RESULT 15
I40884
cytotoxin L - Clostridium sordellii
C:Species: Clostridium sordellii
C>Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 09-Jul-2004
C:Accession: I40884
R:Green, G.A.; Schue, V.; Monteil, H.
Gene 161, 57-61, 1995
A:Title: Cloning and characterization of the cytotoxin L-encoding gene of Clostridium sordellii
A:Reference number: I40884; PMID:95369733; PMID:7642137
A:Accession: I40884
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2364 <RES>
A:Cross-references: UNIPROT:Q46342; UNIPARC:UPI00000802BE; EMBL:X82638; NID:g1000694; PI
C:Keywords: cytotoxin

Query Match 5.0%; Score 358.5; DB 2; Length 2364;
Best Local Similarity 20.3%; Pred. No. 7.6e-10;
Matches 233; Conservative 157; Mismatches 368; Indels 389; Gaps 56;

QY 397 DYDN-----SNPVVQAEQLNHLHYLMNWGD--IVMGDKDANFDG-----VRV 436
Db 1347 DVDVNVKNTIESDEIQGELIENILSKLIEDKNILNNHTINFYDINESNRFLSLTF 1406
QY 437 DAVDVNADLLQIORDYYKAY-----GTQNEKNAIDHLSILEANSNDN 482
Db 1407 SILEDIN---IIEIDLVSYSKILLSGNCMKLIENSDDIQK--IDHI-----GFNGEHQ 1457
QY 483 -----DYVKDQNFSLINDORSGLKAFGVASAYRGNLSNLATAGLNKRSN-----NP 532
Db 1458 KIPIYSYIDNEYKNGFYDYSKEG-----LFTAEFNSIIRNIYMP 1500
QY 533 DSDPVPVYFIRAHSEVQTRIAKIIREKLGKTNADGLNLTDLNKAFFDIYNQDMNAT 592
Db 1501 DS-----NNLFYSSKDLKDIRI-----INKG-----DVKLL 1527
QY 593 DKVYYPNNLPMAYAWMLQNKDVTTRYVYGYDMYTDNGQYMATKTPFPYNAJETLLKGRIKV 652

Search completed: February 11, 2006, 19:42:44
Job time : 40.1742 secs

Db 1528 IGNFKDDMKVLSLFTIEDTWTIKL-----NGVYLDE-----NGVAQIL----- 1566
QY 653 AGQAVSYKQDWSGGILTSVRYGKANSASAGANTETNSGMALL--INNRPNFRAYRNL 710
Db 1567 -----KFMNNAKSALNTS--NSLMNFLESINIKNIP--YNNL 1599
QY 711 TLMGAHKSQAYRPLLLSTKDGIATY--LNSDSVDSRQY-----KYTDSQGNLSF 759
Db 1600 DPNIEFILDNT-----FIISGNSIQGFELICDKDKNIQFYFINFKIKTSYTLTYGN-- 1652
QY 760 SASLEQSVANAQVSGMIQVWPVPGAADNQDVRTSPSTQATKDGNINYHQSDALDSDSVIYEG 819
Db 1653 -----RQNLIVEFSYHLDDSGNI-----SSTVI----- 1675
QY 820 FSNP-QAFAQSPDQYTNNAVIKNGDLFKSWGITQFEMAPQVYSSE-----DGFFLDSVI-L 873
Db 1676 --NFSQKLYGIDRYVNVKVIITAPNLYTDEINITPV-YKPNYICPEVILTDANYINEKINV 1732
QY 874 NGYAFSDRY-----DLAMSKNNKYGSKQDL-----ANAIKQLQSAGIKVLSDLVFNQL- 921
Db 1733 NINDLSIRYVWDDNGSDLILJANGSEEDNQPVKIRFVNVEK-----SDTAADKLUS 1782
QY 922 YNLPGEVVTATRVNQYQAKSGATINKTPYVANTRSYDYOQYGGKFLDQLKLYPRL 981
Db 1783 FNFSDKQDVSVKI-----ISTFSLAAYSDDGFDDY--EFLGYSLDN-DYFYINS 1828
QY 982 FSTKQISTGPKDPSVKJITNNSAKYFN--GSNILGRGAKYVLSGNYK----- 1028
Db 1829 F-----GNMVSGLIYI-NDSLYYFKPKKNLI--TGFTTIDGNKYFYDFPTKSGAASI 1877
QY 1029 --NLADGKLFPTVLNNTYGOPOVSANGFISKNGGIHY-----LDKN--GOEVKNRPK- 1077
Db 1878 GEITIDGKDY----YFNKQGLQV--GVIINTSDGLKYFAPAGTLDENLEGESVNFIGKL 1930
QY 1078 EISGSWYFYFSDGKMATCKTKIGNDTYLFPMPNGKOLKEGVWYDGGKAYYDDNGRTWTK 1137
Db 1931 NIDGKIYFEDNYRAAEVWKLDDDETYFNPKTGEALQHLQIHDNKKYFDDNGIMQT-- 1988
QY 1138 GFVEPRVDGQKWRYFNGDGTIAIGLSLDNRTLYFDAYGYQVKG----- 1182
Db 1989 GFITI----NDKVFFYFNNDGVWQVGYIEVNGKYFYFGKNGERQLGVFNTPDGFKFGPKD 2044
QY 1183 -----QTVINGKSYTFD-----ADQGLVQ-TDNANPAQG--- 1213
Db 2045 DDLGTGEGELTYNGILNFNGKIYFFDISNTAVWGMGTLDGSGTYFYFDNRAEACIGLTV 2104
QY 1214 -----QAGWKLLGDNOMGYRKGQGLLTGEQITIDGQKVFQDNGVQVKG--- 1256
Db 2105 INDKKYFDDNGIRQLGHFITINDNIFYSESQKIELGYQNINGNYFYIDESGLVLIGVFD 2164
QY 1257 -----GTATDASGVLR-----FYDRDQGHQVQKGYSTSDNNWYVN 1293
Db 2165 TPDGKYFAPLNTVNDNIYQAVKYSGLVVRVNEDEVYFGETYKTIETGWIENETDKYFPDP 2224
QY 1294 ESGQVLTGLQITIDGQTVYFDDKGIQAKGKAVDENGNLRYFDADSGNMLRDRKKNVDGNW 1353
Db 2225 ETKKAYKGINVVDDIKYFYDENGIMRTGLISFENN--YYFNE--GKMQFGYLNLIKDKM 2280
QY 1354 YFYNRNG 1360
Db 2281 FYFGKDG 2287

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 18:57:35 ; Search time 184.185 Seconds
(without alignments)
5228.676 Million cell updates/sec

Title: US-10-797-821-39

Perfect score: 7230

Sequence: 1 MEKNLYKLHKVKKQWVAIG.....WKNVDGNWYFNRNGLATRW 1365

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2165443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Uniprot 05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7230	100.0	1365	1	GTFS STRDO
2	5741.5	79.4	1338	2	Q9WJ4 9STRE
3	3172	43.9	1462	1	GTFD STRMU
4	3102.5	42.9	1577	2	Q54178 STRGN
5	3088	42.7	1575	2	Q9LCH3 STROR
6	3080.5	42.6	1554	2	Q8KZL5 9STRE
7	3079.5	42.6	1449	2	Q55264 STRSL
8	3076.5	42.6	1449	2	Q68542 STRSL
9	3052.5	42.2	1597	1	GTFL STRDO
10	3031	41.9	1590	2	Q59983 9STRE
11	3026.5	41.9	1577	2	Q55265 STRSL
12	3023	41.8	1592	1	GTFD STRDO
13	3020	41.8	1455	1	GTFC STRMU
14	3018	41.7	1590	2	Q55263 9STRE
15	2992.5	41.4	1512	2	Q9WJ4 9STRE
16	2992	41.4	1476	1	GTFL STRMU
17	2975.5	41.2	1506	2	Q56CX8 9STRE
18	2942.5	40.7	1454	2	Q69A94 LEUME
19	2891.5	40.0	2835	2	Q8G9Q2 LEUME
20	2863.5	39.6	1599	2	Q00599 STRSL
21	2804	38.8	1518	2	Q00600 STRSL
22	2702	37.4	1508	2	Q52224 LEUME
23	2692	37.2	1508	2	Q9EZHS LEUME
24	2687.5	37.2	1561	2	Q5SBM8 9LACO
25	2656.5	36.7	1477	2	Q9L466 LEUME
26	2639.5	36.5	1330	2	Q84CN4 LEUME
27	2638	36.5	1527	2	Q8KEE1 LEUME
28	2634	36.4	1522	2	Q6TXV4 LEUME
29	2616	36.2	1527	2	Q9ZAR4 LEUME
30	2592.5	35.9	1290	2	Q48756 LEUME
31	2467	34.1	1463	2	Q5SBM6 LACFE

32	2399	33.2	1595	2	Q5SBM3 LACSK	Q5sbm3 lactobacill
33	2247.5	31.1	1772	2	Q5SBN3 LACRE	Q5sbn3 lactobacill
34	2227	30.8	2057	2	Q9RE05 LEUME	Q9re05 leuconostoc
35	2193.5	30.3	1781	2	Q5SBL9 LACRE	Q5sbl9 lactobacill
36	2167	30.0	1772	2	Q5SBN0 LACRE	Q5sbn0 lactobacill
37	2148	29.7	1781	2	Q4UCS4 LACRE	Q4jcs4 lactobacill
38	2146	29.7	1781	2	Q4JLC7 LEUME	Q4jlc7 leuconostoc
39	2065.5	28.6	1016	2	Q9LCJ7 LEUME	Q9lcj7 leuconostoc
40	1344	18.6	1619	2	Q5SBM0 LACRE	Q5sbm0 lactobacill
41	1320	18.3	1231	2	Q5SBN1 LACRE	Q5sbn1 lactobacill
42	1231	17.0	591	2	Q8VUH3 STRMU	Q8vuh3 streptococc
43	923	12.8	522	2	Q8VV10 STRSA	Q8vv10 streptococc
44	670.5	9.3	374	2	Q6ZX19 9LACO	Q6zx19 lactobacill
45	543	7.5	2817	2	Q97K42 CLOAB	Q97k42 clostridium

ALIGNMENTS

RESULT 1

ID	GTFS_STRDO	STANDARD;	PRT;	1365 AA.
AC	R29336;			
DT	01-DEC-1992 (Rel. 24, Created)			
DT	01-DEC-1992 (Rel. 24, Last sequence update)			
DT	13-SEP-2005 (Rel. 48, Last annotation update)			
DE	Glucosyltransferase-S precursor (EC 2.4.1.5) (GTF-S) (Dextranase)			
DE	(Sucrose 6-glucosyltransferase).			
GN	Name=gtfS;			
OS	Streptococcus downei (Streptococcus sobrinus).			
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;			
OC	Streptococcus.			
OX	NCBI_TaxID=1317;			
RN	[1]_TaxID=1317;			
RP	NUCLEOTIDE SEQUENCE (GENOMIC DNA).			
RX	STRAIN=MFE28;			
RX	MEDLINE=90316665; PubMed=2142479;			
RA	Gilmore K.S., Russell R.R., Ferretti J.J.;			
RT	"Analysis of the Streptococcus downei gtfS gene, which specifies a			
RT	glucosyltransferase that synthesizes soluble glucans.";			
RL	Infect. Immun. 58:2452-2458(1990).			
CC	-!- FUNCTION: Production of extracellular glucans, that are thought to			
CC	play a key role in the development of the dental plaque because of			
CC	their ability to adhere to smooth surfaces and mediate the			
CC	aggregation of bacterial cells and food debris			
CC	-!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl)(n) = D-			
CC	fructose + (1,6-alpha-D-glucosyl)(n+1).			
CC	-!- ENZYME REGULATION: Glucan synthesis by GTF-S is independent of			
CC	primer glucan unlike GTF-I.			
CC	-!- MISCELLANEOUS: Synthesizes water-soluble glucans (alpha 1,6-			
CC	glucose).			
CC	-!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.			
CC	-!- SIMILARITY: Contains 10 cell wall binding repeats.			
CC	This Swiss-Prot entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use as long as its content is in no way modified and this statement is not			
CC	removed.			
DR	EMBL; M30943; AAA26898.1; -; Genomic_DNA.			
DR	HSP; P06653; IHG.			
DR	InterPro; IPR002479; Cell wall bd put.			
DR	InterPro; IPR003318; Glyco_hydro_70.			
DR	Pfam; PF01473; CW binding_1; 2.			
DR	Pfam; PF02324; Glyco_hydro_70; 1.			
KW	Dental caries; Glycosyltransferase; Repeat; Signal; Transferase.			
FT	SIGNAL 1 36			
FT	CHAIN 37 1365			
FT	REPEAT 157 197			
FT	REPEAT 178 197			
FT	REPEAT 1062 1082			
FT	REPEAT 1083 1102			

778 VWPVGAADNQDVRTSPSTQATKDNINIHOSDALDSQVIYEGFSNFQAFQSPDQYNAV 837
823 VWPVGAADNQDVRVAASNRKANATQVYESSALDSQLIVEGFSNFQDFVTKSDYTNKK 882
838 IAKNGDLFKSWGTLQPEMAPOYVSSDGTFLDSVILNGYAFSDRYDLAMSKNNKYSQD 897
883 IAKVQLFKSWGVTSPFMAPOYVSSDGTFLDSVILNGYAFSDRYDLAMSKNNKYSQD 942
898 LANAIKGLQAGIKVLSDLPNQLYPLGKEVVTATRVNOYGOAKSGATINTKTPYVANTR 957
943 MINAVKALHKSGLQVIADWPDIYNLPGKEVVTATRVNDYGEYKDSKNTLYAANTK 1002
958 SYG-DYQEQYGGKFLDDLQKLPRLFKSTKQISTGKPIDPSVKITNNSAKYFNGSNILGRG 1016
1003 SNGKDYQAKYGGAFSELAAYPSIFNRTQISNGKKIDPSEKITAMKAKYFNGTNILGRG 1062
1017 AKYVLSF--GKYNLNLADGKFLPTVLNNTYGPQVPSANGFISKNGGIHLDKNGQEVKN 1074
1063 VGVLDKDNASDKYFELKGNQYLPKQMTN-----KEASTGFVNDGNGMTFYSTSGYQAKN 1117
1075 RF-KEISGSMYFSDSKMATGKTKIGNDTYLFMPNGKOLKEGVW--YDGKAYYYDDNG 1131
1118 SFQDAGKNWYFDNNGHMYGLQHLNGEVOYFLSNGVQLRESFLENADGSK-NYFGLHG 1176
1132 RTWNKGFEFRVDGQDKWRYFNGDGTIAIGLSVLDNRTLFPDAYGYQVKGQTVT-INGK 1190
1177 NRYSN-GYYSF--DNDSKWRYFDASGVMAVGLTKTNGTQYFQDGYQVKGAMITGSDGK 1233
1191 SYTFADQDGL-----VQTDNANPAQGGAGKLLG----- 1221
1234 KRYFDDGSGNMAVNFANDKNGDWYLYNSDGIALVGVQTINGKTYFYFGDQKQIKIIT 1293
1222 -----DNQWGY-RKDGQLLTGEQITIDGQKVFQFDNGVQVKG 1256
1294 DNCKLYFLANSSELARNIFATDSQNNWYFSGDGVAVTGSQIIAGKCLYFASDGGQVKG 1353
1257 GTATDASGLRFPYDRDQGHQVGKGYSTSDNNWYVNESQVLTLGLQTDIGQTVYFDDKG 1316
1354 SFVT-YNGKVHYHADSGELQVNRFEADKGNWYLYDSNGEALTGSQRINGQRVFFREG 1412
1317 IQAKGKAVDENGRLYFADSGNMLR-----DRW 1346
1413 KQVKGDAVDYDERGLLRYDKNSGNMYNKKVVTILANGRRIGIDRW 1456

RESULT 4

Q54178_STRGN PRELIMINARY; PRT; 1577 AA.
AC Q54178; Q54247;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Glucosyltransferase.
GN Name=gtfG;
OS Streptococcus gordonii.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OC Streptococcus
OX NCBI_TaxID=1302;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Challis;
RX MEDLINE=96157084; PubMed=8586195;
RA Vickerman M.M., Sulavik M.C., Clewell D.B.;
RT "Molecular analysis of Streptococcus gordonii glucosyltransferase
phase variants.";
RL Dev. Biol. Stand. 85:309-314 (1995).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Challis;
RA Minick P., Vickerman M.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; U12643; AAC3483.1; -; Genomic_DNA.

PIR; B41898; B41898.
DR HSP; P06653; 1HCX.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 5.
DR Pfam; PF02374; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1577 AA; 177805 MW; 5AB0328DC5E08D18 CRC64;
Query Match 42.9%; Score 3102.5; DB 2; Length 1577;
Best Local Similarity 44.9%; Pred. No. 1.2e-143;
Matches 657; Conservative 218; Mismatches 464; Indels 125; Gaps 30;
QY 1 MEKNLYKHLKVKKQWVAIVTTLVLSF-----LAGQVVAADTN----- 41
DB 2 MEKVHYKHKVKQKQWVAIVTSLALLVAPKALGLESGLYVADDAQVTVNKQSASVQSK 61
QY 42 -----NDGTSYQVNMVPSDPK-FDAQAQNGQLAQAMF----- 73
DB 62 DSEQTTSKATDSSQLEKQASSSKETQASATNPANEQTTQDQKEVETSRDSRHE 121
QY 74 -----KANQADQATSQ--VSPATDGRVDNQVTPAANQPAANVAN 112
DB 122 LTQKTSDDSSKSGSQEPKQVADQAESTDQTAALQAKQDSRANDQEBETTENVAKATVSD 181
QY 113 QDVANPATD--AGALNROSAD--TSTDGKAVP---QTSQDPGHLETVDGKTYVVDAN 163
DB 182 KIATPKKRLPEPPAQKESITEKMLAAQAAPVNTHEHDDVLAIHKITIDGKKYVQDD 241
QY 164 GORLKNYSWVIDGKTYFYFDGQGEAQTDLPKTQQAQNDNVPSYQANNOAYSNEASSFET 223
DB 242 GTVKKNFAVELNGKILYFDAETGALVDSNEYQFQGTSSLNNEFTQKNFYGTDDKDIET 301
QY 224 VNYLTADSWYPRKTLKNGQSQWQASSEGDLRPLMTWMPDAAATKAAAYANFNAKEGLISG 283
DB 302 VDGYLTAWSYRPKFLDKGKTWTASTETDLRPLMAWPKDKRTQINLYNMNQENLIG 361
QY 284 SYRQNSAN--LDAATQNIQSAIEKTIASEGNTNWLFDKMSOFVKSQNSIASENETYP 341
DB 362 AFESKTEQVLLTNVQVQVKIEERISKEGDTKWLTLMSAFVKTQPNWNKITESETGT 421
QY 342 NQDHMQGALLFSNSKDTSEHANSWRLNRPFTQKQKIF-TTNVAGYELLANDVDN 400
DB 422 NKDHLGGALLYTNSDKTSHANSRYELNRTPTSGTQPKYFDKNSGGVEFLANDFDN 481
QY 401 SNPVQAEQLNHLHYLMNWDIVMGDKANFDGVRVDAVDNVNADLLQIORDYKAKYGT 460
DB 482 SNPAVQAEQLNHLHYLMNWFSGSIVANDPTANFDGVRVDAVDNVNADLLQIASDYFKRYK 541
QY 461 DQNEKNAIDHLSILEAWSGNDYVQDNFSLISINDORSGLKAFGVSAYRGNLSNL 520
DB 542 GESEBEALKHLISILEAWSNDPDYKQTKGAQLAIDNKLRLSLYSFMKRLSIRSGVEPT 601
QY 521 ATAGLKNRSA-NPDSDPVNPVYFIRAHDSVOTRIAKIIRKLGKTNADGLTNLLDLDN 579
DB 602 ITNSLNDSTRSTENKERTANYIFVRAHDSVOTVIADIIRENI-NPNTDGLT-FTWDELK 659
QY 580 KAFDIYNQDMNATDKVYYPNNLPMAYAMWLQNKDTVTRYVYGDYMDNGOYMATKTPFYN 639
DB 660 QAFKIYNEDMRKADKKYQFNIPHTAHALMSNKDSITRVYVYGDLYTDDQYMEKKSPYHD 719
QY 640 ALETLLKGRIVKVGQ--AVSY-----KQDWS-SGILTSVRYGKANSASDAGNTEFR 690
DB 720 AIDALLRARIKYVAGQDMKVTYMGVPREADKWSYNGILTSVRYGKGANEATDEGTAETR 779
QY 691 NSGMALLNNRPNFR--AYRNLTLNMGAAKQVAPRLLSTKDGITATYLNDSVDVDSROY 748
DB 780 TQGMVAVIASNNENKLNHWDKLVQNVGAHKKQYRVPVLLTTKDGILSRYLTDVEVQSLW 839
QY 749 KYTDSQNLFSASELSQSVANAQVSGMIQVWPVPGAADNQDVRTSPSTQATKDNIIYHOS 808

840 KKTANGILTFDMNDIAGSYNVQSGYLAWVPVGAKEQDARVATSKKKKNGSQGVYESS 899
809 DALDSQVIYEGSNFOAFQSPDQYTNNAIKNAGDLFKSWGITQFEMAPQYVSSSDGTPL 868
900 PALDSQLIYEGSNFODFATRDQYTNKVIKNVNLFEKGVTSFELPPQYVSSQDGTPL 959
869 DSVILNGYAFSDRYDLAMSKNNKYSGKDLANAIKGLQAGIKVLSDLVNPOLYNLPGE 928
960 DSIQNGYAFSDRYDWMKNNKYSGLDOLLNLRALHSVNIQALADWVPDQIYNLPGE 1019
929 VVTATRVNGOYGAQSGATITNKTPYVYANTRSYG-DYQEOYGGKFLDLDLQKLYPLRSTKQI 987
1020 VVTATRVNNGYTYREGABIEKELYVANSKTNGTDYQGYGGAFLDELKAKYBEIPEVQI 1079
988 STGKPIDPVSKITNNKSAKYNFNGSNILGRGAKYVLSB-GNK-YLNLAGDKFLPTVLNNTY 1045
1080 SNGKMTTDEKITKWSAKYFNGTNILGRGAYVYLDWGSKEYLSNKNGETALPKQLVN-- 1137
1046 GQPOYSANGFISKNGGIHYLDKNGQEVKNRP-KEISGSWYVFDSDGKMATGKTIGNDTY 1104
1138 ---KEASTGFVKDTNGFKFYSTSGNOAKDTFIQDENGWNYFDNOGYLVYTGAREIDGKOL 1194
1105 LEWPNKGKQLKEGVWYD-GKKAYYYDDNGRTWTKNGFVEPRVDGQDKWRYFNGDGTIAIGL 1163
1195 YFMKNGVQLRDALQEDENGQYYYDKTGAKVLRYYTS---DQON-WRYFDAKGVWARGL 1250
1164 VSLDNRTLYFDAYGVQVQGTVTI-NGKSYTFDADGDILVQTDNANPAPQGAQGWKLJGD 1222
1251 VKIGDQQYFDQNGYQVKGKVRKQKLYRFDKDSGNVAVINRFAQ-----GD 1298
1223 N--OWGY-RKDGQLLTGEQITDGQKVFQDNGVGVKGGTATDASGVLRFRYDRDQGHQVK 1279
1299 NPSDMYTFYFGDGVALTGLQIKGOOTLYFGQDGQKQVGVVMLADSKSIRYFDANSGBMAVN 1358
1280 GWYSTDDNNWYVNESGQVLTGLQITDGTQVYFDDKGIQAKKAVMDENGLRYFDADSG 1339
1359 KFAEGAKNEWYFDQDGKAVTGLKTINNVLVFDQDGQKQVQVTLADKSIYFDANSG 1418
1340 NMLDRW-KVNDGNWYFNRNGLA 1362
1419 EMAVNFKAEGAKNEWYFDQDGKA 1442

RESULT 5
Q9LCH3 STOR AC Q9LCH3 STOR PRELIMINARY; PRT; 1575 AA.
ID Q9LCH3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Glucosyltransferase.
GN Name=gtfR;
OS Streptococcus oralis.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OC Streptococcus
OX NCBI_TaxID=1303;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC10557;
RX MEDLINE=202311779; PubMed=10768934;
RX DOI=10.1128/IAI.68.5.2475-2483.2000;
RA Fujiwara T., Hoshino T., Ooshima T., Sobue S., Hamada S.;
RT "Purification, characterization, and molecular analysis of the gene
RT encoding glucosyltransferase from Streptococcus oralis.";
RL Infect. Immun. 68:2475-2483(2000).
DR EMBL; AB025228; BAA95201.1; -; Genomic_DNA.
DR HSSP; P06653; 1HCX.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW binding.1; 4_
DR Pfam; PF02324; Glyco_hydro_70; 1.

KW Transferase.
SQ SEQUENCE 1575 AA; 176792 MW; 772A26EAD7C2E543 CRC64;
Query Match 42.7%; Score 3088; DB 2; Length 1575;
Best Local Similarity 44.0%; Pred. No. 6.3e-143;
Matches 669; Conservative 220; Mismatches 456; Indels 176; Gaps 33;
QY 1 MEKNLRYKLHKVKKQWVAIGVTTVLSF-----LAGGVVAADTN----- 40
DB 2 MEKHIYKMHVKYKQWVAIALTTLIVAPKVLGLESVVHADDVKVQVVQBPATAQTSG 61
QY 41 -----NNDGTSVQVKNKVPSPDKPD 60
DB 62 PQQTPPAQKIASQEAQKVTTPADKVTDDVAASEKPAKPAENTEATVQNAQEPKAPADT 121
QY 61 AQAQNGQLAQA-MFKAANQ-----ADQATTSQVSPATDGRVD-NQVTPAANQPAAN 109
DB 122 KEASPEKAAVAEEVKAANAITEIPKTEVADQNKQARPTTAQDQEGDKREKTAVEDKIVAN 181
QY 110 --VANQD-VANPATDAGALNRQSAADTSTDGKAVPQTSQDOP-----GHLETVDGKTYVDA 162
DB 192 PKVAKKDLPEPGSKQGAIAERMVAD---QAQAPVADHDDDLVLSHIKTIIDGKNYYVD 238
QY 163 NGORLKNYSWIDGKTYFFDGTGQTEAQTDLPKTGOANQDNVPSYQANNOAQASNEASSPE 222
DB 239 DGTVKCNFAVELNGRILYFDAETGALVDSNEVFOQGTSSLNNEFSQKNAPFYGTTDKDIE 298
QY 223 TVDNYLTADSWYRPRKILKNGSQWASSEGDLRPLMTWMPDAATAKAAVANFWAKEGLIS 282
DB 299 TVDGYLTADSWYRPRKILKNGSKWTASTETDLRPLMAWMPDKRTQINLYNLMNQQLGA 358
QY 283 GSY--RQNSANLDAATQNIQSAIEKKIASGNTWLRDKMSOFVKSQNQSASENEITY 340
DB 359 GAFENKVEQALLTGASQVQRKIEBKIGEGDKTWLRTLMGAFVKTQPNWNITKTESFTG 418
QY 341 PNODHMOGALLFNSKDETHANSDWRLNLRNPTFTQTKQKYF-TTNVAGYELLANDVD 399
DB 419 TKDHLQGGALLYTNNEKSPHADSFRLLNRTPTSTQTPKYFIDKSGNGYFELLANDFD 478
QY 400 NSNPVQAEQLNHLHYLMNWGDI VMGDKDNFGRVDAVDNVDNADLLQIQRIQRIYKAKYG 459
DB 479 NSNPVQAEQLNHLHYLMNWFGSIVANDPTANPDGRVDAVDNVDNADLLQIASDYFKSRYK 538
QY 460 TDQNEKNAIDHLSIIEANSGNDNDYVQDNFSLIDNDORSGLMKAPGASAYRGNLSN 519
DB 539 VGESEEAIKHLSILEAWSNDPDYNKDTKGQAQLAIDNKLRLSLLYSEFMNLSIRSGVFP 598
QY 520 LATAGLKNRSA-NPDSDPVNPVYFIRAHDSQVOTRIAKIIRKLGKTNADGLTNLTDDL 578
DB 599 TITNSLNRSEKKNGERMANYIFVRAHDSQVOTVIADIIRENI-NPNTDGLT-FTMDEL 656
QY 579 NKAFDIYNQDMNATDKVYPPNNLPMAYAMLQNDQTVTRVYVYDGMVYDNGQYMATKTPPY 638
DB 657 KQAFKIYNEDMRKADKKYTFQNIPTAHALMSNKDSITRVYVYDGLYDQDQYMEKKSYPH 716
QY 639 NAIETLLKGRKYVAGGO--AVSY-----KQDWS-SGLITSVRYCKGANSASDAGNTET 689
DB 717 DAIDALLRARIKYVAGGQDMKYTMGVPREADKWSYNGILTSVRYGTGANEATDEGTAT 776
QY 690 RNSGMALLINNRPNFR--AYRNLTILNMGAHKSQAYRPLLLSTKDGIAIYLANDSDVDSRQ 747
DB 777 FTQGMVAVTASNPNLKLNEWDKLVNMGAAHKNQYRPLLTTKDGISYLLDEEVQSL 836
QY 748 KYKTDSQGNLSFSASELOSVANAQSGMIQVWVPVGAADNDQVRTSPSQATKQGNHYHQ 807
DB 837 WKTTDANGILTFDMDNDIAGYSNVQSVGYLAVWVPVGAQADQARTTASKKNASQGVYES 896
QY 808 SDALDSQVIYEGFSNFQAFQSPDQYTNNAIKNAGDLFKSWGITQFEMAPQYVSSSDGT 867
DB 897 SAALDSQVIYEGFSNFQAFQSPDQYTNNAIKNAGDLFKSWGITQFEMAPQYVSSSDGT 956
QY 868 LDSVILNGYAFSDRYDLAMSKNNKYSGKDLANAIKGLQAGIKVLSDLVNPOLYNLPGE 927

Db 1000 DMFTVMNISTGKPIDSTKIKQWEAKYFNGTNTVLGKAGYVLSDDATGKYFTVNGDFL 1059

Qy 1038 PTVLNNYQPOVSANGFTISKNGGIHYLDKNGOEKVRKEISGWYYPDSGKMATG-- 1095

Db 1060 PA-----SFTGDQNAKTGFYIDGTGMAYYSTGSKAVNSFIYEGGHYYPFDKDHMTGSY 1115

Qy 1096 KTKIGNDTYLPWPNKGQKGEVYDGK-KAYYYDDNGRWTWKG-----FVEFRVDGQDK 1149

Db 1116 KAEDGND-YFPLNGIQMRDAIYQDAQNSYIYGRTGILY--KGNWYFPFVDPNNANKTV 1172

Qy 1150 WRYFNGDGTIAIGLSLNDRLTYFAYGQVKGQVTINGKSYTFDADQG----- 1199

Db 1173 FRYEDANNVMAIGYRMYGQTYFEBENGFAQKQLLTDKDGTHYFDEONGAMAKNFVNV 1232

Qy 1200 --DLVQTDNANPAQOQ----- 1214

Db 1233 GDDWYMYMDGNAGVAKGQYVWNQILYFNPETGVQVKGQFITDAQRTSYDANSALKSS 1292

Qy 1215 -----AGWKLLGDN--ONGY--RKDGQLLTGEQITIDGQKVFFQ-DN 1250

Db 1293 GFPTNGSDMYAENGYYVYKGFQVAENODQWYFDTTGKQAKAAKVDGRDLVFNPD 1352

Qy 1251 GVOVKQGTATDASGLRFLFYDRDQGHQVGKWTSTSDNNWVYVNESQVLTGLQITD---- 1306

Db 1353 GVOVKGDFATDSGNTSFVHGDNQDKVVGFFTTGNNAWYADNNGNLVKGFQEIQGWY 1412

Qy 1307 -----GQTVYFD-DKGIQAKGKAVDENGRLRYFDDADSGNMLDRWKN 1348

Db 1413 HFDEVTGQAKGAALVNGQQLYFVDSGTVQVKGDFVTDQGNYSYDVNSGD-----KK 1466

Qy 1349 VDG-----NWYFNRNG 1360

Db 1467 VNGFTTGDNWYADGQG 1485

RESULT 7

OS5264_STREL PRELIMINARY; PRT; 1449 AA.

AC OS5264;

DT 01-NOV-1996 (TremBLrel. 01, Created)

DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)

DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)

DE Glucosyltransferase precursor.

GN Name=gfll;

OS Streptococcus salivarius.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI_TaxID=1304;

RN [1]_TaxID=1304;

RP NUCLEOTIDE SEQUENCE.

RX MEDLINE=95122197; PubMed=7822030;

RA Simpson C.I., Giffard P.M., Jacques N.A.;

RT "Streptococcus salivarius ATCC 25975 possesses at least two genes

RT coding for primer-independent glucosyltransferases.";

RL Infect. Immun. 63:609-621(1995).

DR EMBL; J35495; AAC41412.1; -; Genomic_DNA.

DR PIR; T30857; T30857.

DR GO; GO:0016740; F:transferase activity; IEA.

DR GO; GO:0009250; P:glucan biosynthesis; IEA.

DR InterPro; IPR002479; CW.binding.

DR InterPro; IPR003318; Glyco_hydro_70.

DR Pfam; PF01473; CW.binding_1; 2.

DR Pfam; PF02324; Glyco_hydro_70; 1.

KW Signal; Transferase.

FT SIGNAL

FT CHAIN 1 35 Potential.

FT CHAIN 36 1449 Glucosyltransferase.

SQ SEQUENCE 1449 AA; 159984 MW; DD62F07306B586A46 CRC64;

Query Match 42.6%; Score 3079.5; DB 2; Length 1449;

Best Local Similarity 45.0%; Pred. No. 1.5e-142;

Matches 667; Conservative 206; Mismatches 451; Indels 157; Gaps 35;

Qy 1 MEKNRLRYKLHKVKQWVAIGVTTVTLSPLAGQVVAADTNNNDGTSVQVKNKVPSPDKFD 60

Db 1 MDKKVHYKHKVKQWVAIGVTTVTLSPLAGQVVAADTNNNDGTSVQVKNKVPSPDKFD 48

Qy 61 AQAQNGLOAQAMFKAANO--ADQATATSOVSPATDGRVDNQVTPAA-----NOPAAN 109

Db 49 ATVAIPDITVDGTGVSNDTAAQDPTTAAATNDVATD-QATPTATFDLTTDTTNTVAAN 107

Qy 110 VANQDVANPATDAGALNRQSA-----DTSTDGKAV----- 140

Db 108 -AVDTVATVGTDRRAATTNDTATNDTAVDTTNNNTTTTDTTVDRAATTERRATGARROP 166

Qy 141 -----PQT-----SDQGHLETVDGKTVYVDANGORLKN 169

Db 167 TCGRRATPVNGTNNWNTVTVVNDLPATNNVTDGFSHIKTINGKQYVYEDDGTIRKN 226

Qy 170 YSMVJDKTYYPDGQTGEAQT-----LPKTOANQDNVPSYOAN-----NOAYSNEA 218

Db 227 VYLERIGSGQYFNAETGELSNOKEYRFDKNGGTGSS--ADSTNTNTVTVNGDKNAFVGTTD 284

Qy 219 SSFETVDNYLTADSHYRPRKILKNGQSWQASSEGDLRFLMTWTPDAATKAAYANFWAKE 278

Db 285 KDIELVDGYFTANTWYRPEKILKDGKEWTASTENDKRPFLTVMWFSKAIQASLYLNMKEQ 344

Qy 279 GL-IGSYRONS--NLDAATQNIQSAIEKKIASEGNTNWLDRKMSQFVKSONQWSIAS 335

Db 345 GLGTTNQTYSSTQTMQDAALEVQKRIEERLAREGNTDMLRTTIKNFVKTPQGHNSISE 404

Qy 336 NETVYPNQDHMQGALLFSNSKOTEHANSWRLLNRNPTFQTKQ--KYF-TTNYAGVEL 392

Db 405 N---LDNNDLHQQGALLYNNDSRTSHANSYRLLNRTPTSTQTKENPKYTKDTSNGPPEF 461

Qy 393 LLANDVDSNPVVOAEQLNHLHYLMNWGDIVNGDKDANFDGVRVDAVDNVDNADLLQIQRD 452

Db 462 LLANDIDNSPAVQAEQLNWLHYIMNIITIGGSEDEFDGVVRVDAVDNVDNADLLQIASD 521

Qy 453 YVKAKYGTDONEKNAIDHL-SILEAMSGMNDNVVQDNNFSLSDNDQSGMLKAFGYASA 512

Db 522 YFKAKYGADQSQDAIKHLSILEAMSHNDAYNEDTKGAQLPMDPMLHALYLSLLRPTG 581

Qy 513 YRGNLSNLATAGLKNRS--ANPDSVPVNYFIRAHDSVQTRIAKIIREKLGKTNADGUT 571

Db 582 NRSQVEPLISNLNDRSESGKNSKRMANYAFVRAHDSVQSIIQIINKNEINPQSTG--N 639

Qy 572 NLTLDDLNAKAFDIYNQDNATDKVYYPNNLPMAYAMLQNKDQTVTRVYVYGDYNDNGQYM 631

Db 640 TFTLDEMKKAFIYNKWRANKQYTNIPISAYALMLTKDQTVPRVYVYGDYNDNGQYM 699

Qy 632 ATKTPFYNAIETLLKGRIKYVAGGQ--AVSY-----KQDW-SSGILTSVRYGKGSASD 683

Db 700 AKQSPYDAIETLLKGRIRYAAAGGQDMKNVYIGYNTNGWDAGVLTSVRYGTGANSASD 759

Qy 684 AGNTRNSGMALLINRPNFRAYRNLTLNMGAAHKSQAYRPLLLSTKDGATYLNDSV 743

Db 760 TGTASTRQGMAVIVSNQPALRLTSLNLTINMGAAHRNQAYRPLLLTNDGVATYLNDSDA 819

Qy 744 DSRQKYTDSQGNLSFSASELSQVANAQVSGMIQWVPVGAADNODVRTSPSTQATKDN 803

Db 820 NG-IVKYTDTGNGNLTFSSANEIRGINPQVDGYLAWVPVPGASENQDVRVAPSKENSSGL 878

Qy 804 IYHOSDALDSOVIYEGFSNFQAFQSPQYTNNAVIAKNGDLFPKSWGITQFENAPQVVSSE 863

Db 879 VYESNAALDSQVIYEGFSNFQFVQNPQYTNKKIAENANLFPKSWGITSFEPAPQVSSD 938

Qy 864 DGTFLDSVILNGYAFSDRYDLAMSNNKYGSKQDLANAIKGLQSAGIKVLSLDPVNLQYN 923

Db 939 DGSFLDSVIQGYAFTDRYDIGMSKDKYGLADLKAALKSLHAVGISAIADVPDQIYN 998

Qy 924 LPQKEVVTATRVNQYQAKSGATINKTPYVANTRSYG-DYQBYQCGKFLDLDLQKLYPLRF 982

Db 999 LPQGEVVTATRVNNYGETKDGAIIDHSLYAAKTRTFGNDYQGYGAGFLDELKRLYPPQIF 1058

Qy 983 STKQISTGCKPDPSPVKITNWSAKYPNGSNILGRGAKYVLSSEG-NKYLNLADGKLFPTVL 1041

Db 1059 DRVQISTGKRMWTTDEKITQWSAKYMNQNTNILDNRGSEVVLKNGLNGYVGTNGGKVSPLPKVV 1118
Qy 1042 NNTYGOQVVSANG-----FISKNGIHYLDKNGQEVNRF-KBISGSWYFFSDGKMATG 1095
Db 1119 GSNQSTGNDGNGDGGSGKFKELFSVRYRYNNGQYAKNAFIKNDGNGVYFFDSGRWAVG 1178
Qy 1096 KTKIGNDTYLFMPNGKQKLGWYDVGK-KAYYYDDNGRTWTN-----KGFVE 1141
Db 1179 EKTIDGKQYFFFLANGVQLRDGYRQNRGQVFFYDQNGVLNANGKQDPKPDNNNNSGRNQ 1238
Qy 1142 FRVDGDKWRYFNGDGTIAIGLVSLDNRTLYFDAYGVQVKGQVTTNGKSYTFDADQGD 1201
Db 1239 FVQIGNVWAYDNGKRVTHQNINGQELFFDNNGVQVKGRTVNGAIRYDANSSEM 1298
Qy 1202 VOTDNANPAPQOAGWKLLGDNQWY-RKDGQLLTGEGTTIDGQKVPFQDNGVQVKGGTAT 1260
Db 1299 ARNFAEIEP-----GVWAYFNNDGTAVKGSQNINGQDLYFDQNGRQVKGALA- 1346
Qy 1261 DASGLVRFYDRDQHQVKGKQWYSTDNDNVYVNESQVLTLGLOTIDGQTVYFDKGIQAK 1320
Db 1347 NVDGNLRYDYVNSG-ELYRNRPEHIDGSWYTFDNGNAVKGMVNINGQNLFFDNGKQIK 1405
Qy 1321 GKAVWDENGLRYFDADSGNMLDRWKVD-GNMYFENRNG 1360
Db 1406 GHLV-RVNGVVRYPDENGEMAVNRVVEVSPGWWVYFDGEG 1445
RESULT 8
O68542_STRSL PRELIMINARY; PRT; 1449 AA.
ID O68542; 1449 AA; 159895 MW; 0700F6D748471BFB CRC64;
AC O68542; 1449 AA; 159895 MW; 0700F6D748471BFB CRC64;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Glucosyltransferase N (Fragment).
GN Name=gtfN;
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=V1477;
RA Jaffe R.I.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF049609; AAC05156.1; -; Genomic_DNA.
DR PIR; T30552; T30552.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
DR KW Transferase.
FT NON_TER 1449 1449
SQ SEQUENCE 1449 AA; 159895 MW; 0700F6D748471BFB CRC64;
Query Match 42.6%; Score 3076.5; DB 2; Length 1449;
Best Local Similarity 45.0%; Pred. No. 2.1e-142;
Matches 667; Conservative 205; Mismatches 452; Indels 157; Gaps 35;
Qy 1 MBKNRYKLHKVKKQWVAIGVTTVTLFLAGGQVVAADTNNDGTSVQVKNMVPSPKFD 60
Db 1 MDKKVHYKHKVKKQWVAIGVTTVTLFLAGGQVVAADTNNDGTSVQVKNMVPSPKFD 60
Qy 61 AQANQQLAAQAFKAANQ--ADQATTSQVSPATDGRVDNQVTPAA-----NOPAAN 109
Db 49 ATVAIPDITVDRTGTVSNDTTAAQDPTTAAVATNDVATD-QATPTATFDLTDTTNTVAAN 107
Qy 110 VANQDVANPATDAGALNRQSA-----DTSTDGKAV----- 140
Db 108 -AVDTVATVGTDRAAATNTDTATNDTAVDTTNNNTTDTTNTNRAATERRATGARRGP 166

Qy 141 -----PQT-----SDQPGHLETVDGKTYVVDANGQRLKN 169
Db 167 TGGRRATPVNGTNNANNTVVVNDLPATNNVVTDGPHIKTINGKQYVVEDDGTIRKN 226
Qy 170 YSNVIDGKTYVVDGQTEAQT-----LPTGQANQDVPDSYOAN-----NQAYSNEA 218
Db 227 YVLERIGGSOYFNAETGELSNOKEYRFDKNGGTGSS--ADSTNTNTVNGDKNAFTGTTD 284
Qy 219 SPSFETVNTLTADSWRPRKILKNGOSWQASSBGDLRPIILMTWMPDAATKAAAYANFWAKE 278
Db 285 KDIELVDGFTANTWRPKELLKDGKEWTASTENDRPLLTWMPKAAQASLYNTMKEQ 344
Qy 279 GL-IGSYRQNSA--NLDAATQNIQSAIEKKIASEGNTWLDRKMSQFVKSQNQWQSIASE 335
Db 345 GLGTNTQYTSFSSQTMQDALEVKRIEGRAREGNTDWRITTIKNFVKTPQGMNSTSE 404
Qy 336 NETVYNQDHMOGGALLPSNSKQTEHANSMDWLLNRNPTFQTKQ--KYP-TTNYAGYEL 392
Db 405 N-----LDNNDHLOQGALLYNNDSRTSHANSDYRLNRTPT7SQTGKHPKYTKDTSNGGF 461
Qy 393 LLANDVDNPNVQABQLNHLHYLMNWGDIVMGDKDFVGRVDAVDNVDNADLLQIORD 452
Db 462 LLANDIDNSNPAVQABQLNHLHYIMNIGITGSEDEDFGVRVDAVDNVDNADLLQIORD 521
Qy 453 YPKAKYGTQDNEKNAIDHLSILEAMSGNDNDYVQDNFSLSIDNDQSGMLKAFGYASA 512
Db 522 YPKAKYGTQDNEKNAIDHLSILEAMSGNDNDYVQDNFSLSIDNDQSGMLKAFGYASA 512
Qy 513 YRGNLSLATAGLKNRS--ANPDSDPVNTVPTVTRAHDSQVQTRIAKILIRKLGKTNADGLT 571
Db 582 NRSQVEPLISNLSNDRSGSKSKRMANTAFVRAHDSQVQTRIAKILIRKLGKTNADGLT 639
Qy 572 NLTLDDLKNAFDIYNQDMNATKVVYPNNLPMAYAMLONKDTVTVYVYGDWYTDNGQYM 631
Db 640 TTTDEMKAFAEYINKDMRSANKQYQYINPSAYALMLTHKDTVPVYVYGDWYTDNGQYM 699
Qy 632 ATKTPFYNAIETLLKRIKYVAGGQ--AVSY-----KODW--SSGILTSVRYKGANASASD 683
Db 700 AQKSPYDAIETLLKRIKYVAGGQ--AVSY-----KODW--SSGILTSVRYKGANASASD 759
Qy 684 AGNTETRNSGSMALLNNRPNFRAYRNLTLNMGAAHKSQAYRPLLLSTKDGATYLNDSOV 743
Db 760 TGTAEATRNQGMVIVSNQPALRLTSLNTINMGAAHRSQAYRPLLLSTKDGATYLNDSOV 819
Qy 744 DSRQVKTDSQCNLSFSASELQSVANAQVSMQVWVPGVGAADNDQDVRTSPSQATKDN 803
Db 820 NG-IVKTDGNGNLTFSAEIRNPQVDGYLAVVWVPGVGAADNDQDVRTSPSQATKDN 878
Qy 804 IYHQSALDSDQVIYEGFSNFQAFQSPDQYTNVIAKNGDLFKSWGITQFEMAPQVVSSE 863
Db 879 VYESNAALDSQVIYEGFSNFQAFQSPDQYTNVIAKNGDLFKSWGITQFEMAPQVVSSE 938
Qy 864 DGTFLDSVLNGYAFSDRYDLAMSKNKKYGSQDLANAIKGLQSAQIKVLSLDPNQLYN 923
Db 939 DGSFLDSVLNGYAFSDRYDLAMSKNKKYGSQDLANAIKGLQSAQIKVLSLDPNQLYN 998
Qy 924 LPGAQVVTATRVNQYQKAGSGATINKTPVYVANTSVG-DYQEQYGGKFLDDQLKLYPLRF 982
Db 999 LPGAQVVTATRVNQYQKAGSGATINKTPVYVANTSVG-DYQEQYGGKFLDDQLKLYPLRF 1058
Qy 983 STKQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKYVLSSEG-NKYLNLNADGKLFPTVL 1041
Db 1059 DRVQISTGKRMWTTDEKITQWSAKYMNQNTNILDNRGSEVVLKNGLNGYVGTNGGKVSPLPKVV 1118
Qy 1042 NNTYGOQVVSANG-----FISKNGIHYLDKNGQEVNRF-KBISGSWYFFSDGKMATG 1095
Db 1119 GSNQSTGNDGNGDGGSGKFKELFSVRYRYNNGQYAKNAFIKNDGNGVYFFDSGRWAVG 1178
Qy 1096 KTKIGNDTYLFMPNGKQKLGWYDVGK-KAYYYDDNGRTWTN-----KGFVE 1141
Db 1179 EKTIDGKQYFFFLANGVQLRDGYRQNRGQVFFYDQNGVLNANGKQDPKPDNNNNSGRNQ 1238
Qy 1142 FRVDGDKWRYFNGDGTIAIGLVSLDNRTLYFDAYGVQVKGQVTTNGKSYTFDADQGD 1201

```
Db 1239 FVQIGNWYWDGNGKRVIGHQNINGQELFFDNNGVQKGTVENGAIRYDANSGBM 1298
Qy 1202 VQTDNANPAQOAGWKLGDNQWGY-RXDGQLLTGEBQIDGQKVFQDNGVQVKGGTAT 1260
Db 1299 ARNRPAEIEP-----GWAYFNNDGTAVKGSQINQDLYFDQNGRQVKGALA- 1346
Qy 1261 DASGVLRFRDROCHQVKGWSTSDDNWVYNESQVLTGLQTDGQTVYFDDKGIQAK 1320
Db 1347 NYDGNLRYDYVNSG-ELYENRHEIDGSWYFDGNGNAVKGVMVINGQNLFFDNNGKQIK 1405
Qy 1321 GRAVMDENGLRYFDADSGNMLDRKQVD-GNWYFENRNG 1360
Db 1406 GHLV-RVNGVRYFDPSNGEMAVNRWVEVSPGWVYFDGEG 1445

RESULT 9
GTF1_STRDO
ID_GTF1_STRDO STANDARD; PRT; 1597 AA.
AC F11001;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
GN Name=gtfi;
OS Streptococcus downei (Streptococcus sobrinus).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1317;
RN [1]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC MEDLINE=87308014; PubMed=3040686;
RX Ferretti J.J., Gilpin M.L., Russell R.R.B.;
RA "Nucleotide sequence of a glucosyltransferase gene from Streptococcus
RT sobrinus MFe28.";
RL J. Bacteriol. 169:4271-4278 (1987).
CC -!- FUNCTION: Production of extracellular glucans, that are thought to
CC play a key role in the development of the dental plaque because of
CC their ability to adhere to smooth surfaces and mediate the
CC aggregation of bacterial cells and food debris.
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-
CC fructose + (1,6-alpha-D-glucosyl) (n+1).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes
CC water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
CC forms of glucans.
CC -!- SIMILARITY: Belongs to the glucosyl hydrolase 70 family.
CC -!- SIMILARITY: Contains 19 cell wall binding repeats.
CC
CC This Swiss-Prot entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC
CC EMBL; M17391; AAC63063.1; -; Genomic DNA.
DR InterPro; IPR002479; Cell wall bd put.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 4.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Dental caries; Glycosyltransferase; Repeat; Signal; Transferase.
FT SIGNAL 1 38 Potential.
FT CHAIN 39 1597 Glucosyltransferase-I.
FT REPEAT 1099 1132 A repeat.
FT REPEAT 1163 1213 AC repeat.
FT REPEAT 1227 1277 AC repeat.
FT REPEAT 1292 1342 AC repeat.
FT REPEAT 1352 1399 B repeat.
FT REPEAT 1405 1455 AC repeat.
FT REPEAT 1465 1512 B repeat.
```

```
FT REPEAT 1519 1568 AC repeat.
FT REPEAT 1582 1597 A repeat (incomplete).
FT REGION 39 1050 Catalytic (approximate).
FT REGION 1099 1597 1.25 A, 2 B and 5 AC repeats.
FT REGION 1099 1597 Glucan-binding (approximate).
SQ SEQUENCE 1597 AA; 177080 MW; B9E86A200868798E CRC64;

Query Match 42.2%; Score 3052.5; DB 1; Length 1597;
Best Local Similarity 42.4%; Pred. No. 3.6e-141;
Matches 688; Conservative 210; Mismatches 433; Indels 291; Gaps 40;

Qy 1 MEKNLRYKHLKVKQWVAIGVTIVT-LSFLAGQGVVAAADTNNNDGTSVQVNNKMWSPDKPF 59
Db 1 MEKNERFQHKVKQKQWVTISVASATMLASALGASVASADTETVSEDSQA--VLTAAD--- 55

Qy 60 DAQAQNGQLAQAMFXAANQADQTATSQVSPATDGRVD---NOVTPAANQPAANVANQDV 115
Db 56 --QTTTNQDTEQTSVAA-----TATSEQASATDAATDQASATDQASAAEQTGTTASTDT 108

Qy 116 -----ANPA-----TDAGALNRQSAADSTDGKAVPQTSQDPGHLETVDGK 156
Db 109 AAQTTTANAEAKWVPTENENQVFTDEMLAEAKNVA--TAESNSIFSDLAKMSNVQVDGK 166

Qy 157 TYVVDANGORLKNYSWVIDGKTYYPDGOTG---EAOTDLPKTKGOANQDNVPSQOANNO 212
Db 167 YYYVDQDGNVKKNFAVSVEKIIYFD-ETGAYKOTSKVEADKSG-SDISKEETTFANNR 224

Qy 213 AYSNEASFETVDNYLTADSWYRPRKILKNGSQWASSEGDLRPIILMTWMPDAATAKAA 272
Db 225 AYVSTAENFEAIDNYLTADSWYRPRKILKNGTWTSESSKDDPRILLMAWPDTEKRV 284

Qy 273 NPWAKEGLISGSY---RONSANLDAATQNIQSAIEKKIASEGTNWLRDKMSOFVKSQNW 330
Db 285 NYMKVVGIDKTYTAETSOADLTAAAEVQARIEQKITTEQNTKWLREIAISAFVKTQPM 344

Qy 331 STASENETVYPNODHMOGGLLFNSKD-TEHANSDRLLNRNPTFTGK-QKYFTTN-- 386
Db 345 NGESEK-----PYDDHLQNGALKFDNQSDLTPTQSNRYLLNRNPTNGSLDSRTYAN 400

Qy 387 --YAGYELLANDVNSNPVQAEQLNHLHYLMNMGDI VMGDKDANFGVRVDAVDNVA 444
Db 401 DPLGGYELLANDVNSNPVQAEQLNHLHYLNFGTIVAKDADANFDSIRVDAVDNVA 460

Qy 445 DLLQIQRDIYKAKYGTQDNEKVAIDHLSILEAWSGNDNDYVKDQNNFSLINDORSML 504
Db 461 DLLQISSDYLKAAAYGIDKNNKNNHVSIVEAWSNDTPYLDHDDGDNLMNMNPKRFLSML 520

Qy 505 KAFGYASAVRGNLSNLATAGLKNRSANP-DSDPVPNVYFIRAHDSVOTRIAKITREKLG 563
Db 521 WSLAKPLDKRSLNPLIINSLVDREVDETVSFYSFARAHDSVQDLIRDIKAEI- 579

Qy 564 KTNADGLTNLTLDLNLKAFDIYNQDMNATDKVYYPNNLPMAYMLOWNKDVTTRVYVGM 623
Db 580 NENAFGY-SFTQDEIDQAFKIVNEDLKTKDKYTHYNVPLSYTLLLTNKGSIPTVYVGM 638

Qy 624 YTDNGQYMATKTPFFNAIETLLKGRIVKAGQAVSYKQDWSGSLTTSYRYKGANSAD 683
Db 639 FTDDQYQWANKTVNYDAIESLLKARMKYVAGQAMQNYQIGNGEILTTSYRYKGALQKSD 698

Qy 684 AGNTEFRSGMALLNNRPNRA-VRLNLTNNGAAHKSQAYRPLLLSTKDGATYLNDS 742
Db 699 KGDATRTSGVGVVMGNQFNFLSDGKVALNNGAAHQAHEQYRALWSTKDGATYATDAD 758

Qy 743 VD-SRQYKYSQGNLSFSASBLOSQVANAQSGMIQVWVPVGAADNQDVRTSPSTQATKD 801
Db 759 ASKAGLVKRTDENGYLIFLNDLKGAVNPQVSGFLQWVPVGAADDQDIRVAASDTASTD 818

Qy 802 GNIYQSQALDSQVITYEGFSNFOAFAQSPDQITNAVIKNGDLPKSWGITQFEMAPQYVS 861
Db 819 GKSLLHQDAAMD SRVMFEGFSNFQSFATKEEYTNVVIANNVDKFSWGITDFEMAPQYVS 878

Qy 862 SEDGTFLDSVILNGVAFSDRYDLANSKKNYKSKQDLANAIKGLSAGIKVLSDLVPNOL 921
```

Db 879 STDQGFSLDSVIQNGYAFTDRYDILGMSKANKYGTADQLVKAIKALHAKGLKVMADWVDDQM 938
QY 922 YNLPGKEVTVATRVNOYQAKSGATINKTPYVANTRSYG-DYOEQYGGKFLDLQKLYPR 980
Db 939 YTFPKQEVTVTRTDKFGKPIAGSIQNHSLYVTDTKSSGGDYQAKYGGAFDELDELKEKYPE 998
QY 981 LFTSKQISTGKPTDPSVKITNWSAKYPNGSNILGRGAKYVLSB--GNKYLNLAGDKFLFLP 1038
Db 999 LFTKQISTGQAIDPSVKIKOWSAKYFNGSNILGRGADYVLSDAQSNKYNLNSDDKFLFLP 1058
QY 1039 TVL----- 1041
Db 1059 KILLQGVESGIRPDTGTYVNSSTGCKVTDSEITAGNLVYFGQGVNMTGAQNIKGS 1118
QY 1042 -----NNTYG-----OPQVSANGFISK 1058
Db 1119 NYVFLANGAALRNVTYTDAGQNHYYGNDGKRYENGQQFGNDGSRWYFKNGVWALGLTIV 1178
QY 1059 NGGIHYLDK-----NGQEVKNRP-KEISGSWYTFDSGKMAT 1094
Db 1179 DGHVQVFDKGVQAKDKIIIVTRDGKRVYFDQHNAGVNTTFVADKTHGWYLLGKGVAVT 1238
QY 1095 GKTKIGNDTYLEWPNKGKOLKEGVYDQK--KAYYD-DNGRTWTKGFVEFRVGDQKWR 1151
Db 1239 GAQTVGKQHLVFEANGQQVK-GDFTAKOGKLYFYDVSDDMTNT-FIE---DKAGNWF 1293
QY 1152 YFNGDGTIAIGVSLDNRLTYFDAYGVQVKG-----QVTIN 1188
Db 1294 YLKGCGAATVGAQTKIGKQKLYFKANGQQVKGDIVKDDGKIRYYDAQTGEQVFNKSVSN 1353
QY 1189 GKSYYTADQDGLVQTDNANPAQ-----GQ-----AGWLLGDNONGYRK 1229
Db 1354 GKTYFEGSD--GTAQT-QANPKQTFKDGSGVLRFYNLGQYVSGSGWYETAHEHWYVK 1410
QY 1230 DQGLLTGEOTIDQKVFQDNGVQVKG----- 1256
Db 1411 SGKVLTAQOTIGNQRYFKDNGHQVKGQLVTGNDGKLRYYDANSQDAFNKSVTVNGKTY 1470
QY 1257 -----GTA-----TDASGVLFYDRDQGHQVKGWYSTSDNNWVYNESQVLT 1300
Db 1471 YFGSDTAQTAQNPQKQTFKDGSGVLRFYNLGQYVSGSGWYKQAGQWLYVKD-GKVLT 1529
QY 1301 GLQTDGQTVYFDKGIQAKGKAVDENGNLRYFDADSGMULDRWKNVYDGNWYFNRG 1360
Db 1530 GLQTVGNQKVFYFDKNGIQAKGKAVRTSDGKRVYFDENSGSMITNQWKFYGVQYVYFGSDG 1589
QY 1361 LA 1362
Db 1590 AA 1591

RESULT 10

Q5983 STRE PRELIMINARY; PRT; 1590 AA.
AC Q5983;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5).
GN Name=gtfi;
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1310;
RN [1]_TaxID=1310;
RP NUCLEOTIDE SEQUENCE.
RX STRAIN=OM2176;
RX MEDLINE=94146405; PubMed=8312602;
RA Sato S., Inoue M., Handa N., Aizawa Y., Isobe Y., Katayama T.;
RT "DNA sequence of the glucosyltransferase gene of serotype d
Streptococcus sobrinus.";
RL DNA Seq. 4:19-27(1993).
RN [2]

RP PROTEIN SEQUENCE.
RX MEDLINE=91224988; PubMed=1827439;
RA Mooser G., Hefta S.A., Paxton R.J., Shively J.E., Lee T.D.;
RT "Isolation and sequence of an active-site peptide containing a
catalytic aspartic acid from two Streptococcus sobrinus alpha-
glucosyltransferases.";
RL J. Biol. Chem. 266:8916-8922(1991).
DR EMBL; D13858; BAA02976.1; -; Genomic_DNA.
DR PIR; A39841; A39841.
DR HSP; P06653; LHXC.
DR GO; GO:0047849; F.dextranucrase activity; IEA.
DR GO; GO:0016757; F.transferase activity, transferring glucosyl. . .; IEA.
DR GO; GO:0009250; P.glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW.binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW.binding_1; 2_70.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Glycosyltransferase; Signal; Transferase.
FT SIGNAL 1 38 Potential.
FT CHAIN 39 1590 Glucosyltransferase-I.
SQ SEQUENCE 1590 AA; 175956 MW; C3C83A57CF3C2B0E CRC64;

Query Match 41.9%; Score 3031; DB 2; Length 1590;
Best Local Similarity 41.6%; Pred. No. 4.1e-140;
Matches 672; Conservative 217; Mismatches 436; Indels 292; Gaps 36;

QY 1 MEKNRYKLHKVKQKOWVAIGVTVT-LSFLLAGQVVAADTNNDGTSVQVNMKVPSPDKF 59
Db 1 MEKNVRFQHKVKRWVLSVASATWLASALGASVASADTDTASDDSNQA--VVTGD--- 55
QY 60 DAQAQNGQLAQMFAKAAQADQATATSVSPATDGRVDNQVTPAANQVPAANQV--- 115
Db 56 --QTTNNQATDQTSIAA-----TATSEQASATDAATDQ--ASAAEQTQGTASTDTAAQT 106
QY 116 ---ANPA-----TDAGALNRQSAADTSTDGKAVPQTSDDPGHLETVGDKTYV 160
Db 107 TTNAMEAKVPTTENENQGTDEMLAEAKNVATAESD--SIPDLAKMNVKVDGKYV 164
QY 161 DANGQRLKNYSVIVDKTYVFDQGTG---EAQTDLPKQANQVNDVPDSYQANNQAYSN 216
Db 165 DQGNVKNFPAVSGDKIYFD-ETGAYKDTSEKVDADKSSAVSQA-TIPANNRAYST 222
QY 217 EASSFETVDNYLTADSWYRPRKILKNQSQWASSEGDLRPIILMTWMPDAAATKAAYANFWA 276
Db 223 SAKNFEADVNYLTADSWYRPRKILKDGKTWTSKDDPRFLMAMWPDTEKRYVNYVN 282
QY 277 KEGLISGSY--RQNSANLDAATQNTQSAIEKKIASEGNTWLRDKMSQVFKSQNWSIAS 334
Db 283 KYVGIDKTYTAETSOADLTAAAEVQARIEQKITSENTKWLREAIASFVKTQPOWNGES 342
QY 335 ENETVYVQDQHMGGALLFNSKD--TEHANSQWRLNLRNPTFTQTK-QKYFTTN----YA 388
Db 343 EK-----PYDDHLQNGALLFDNQTDLTPTDQSYNRLNLRNPTFTQTKSLDSRFTYNDPLG 398
QY 389 GYELLANDVNSNFPVQAEQLNHLHLYLMNWGDI VMGDKDANFDGVYRVDVAVNNADLLQ 448
Db 399 GYDFELLANDVNSNFPVQAEQLNHLHLYLMNWGDI VMGDKDANFDGVYRVDVAVNNADLLQ 458
QY 449 IORDYKAKYGTQDQNEKNAIDHLSILEAWGNDNDYVDQNNFSLSIDNDQSGMLKAFG 508
Db 459 ISSDYLKAAAGIDKNNKNNHVSIVEAWSNDTPTLHDDGDNLMNMKDFLSLWLSLA 518
QY 509 YASAVRGNLSNLATAGLKNRANP--DSDPVNVPVFRADHDSVQVTRIAKIIIEKLGKTNW 567
Db 519 KPLDKRSGNLNLIHNSLVDRREVDTVPSTSFARAHDSVQDIIRDIKAEI-NPNS 577
QY 568 DGLTNLTLDLDDKAFDIYNQDMNATDKVYPPNPLPMAYAMWLQNKDVTVRVYGDMDYDN 627
Db 578 FGY-SFTQEEIEQAFKIYNEDLKKTDKKYTHYVPLSYLLLTNKGSIIPRVYGDMDTDD 636
QY 628 GOYMATKTPFYNAIETLLKGRIKYVAGQAVSYKQDMSSGILTSVRYGKGANASDAGNT 687
Db 637 GOYMAKNTVNYDAIESLLKARMKYVSGQAMQNYQTNGEILTSVRYGKALKQSDKGA 696

```
QY 688 ETRNSGSMALLINRRPNERA-YRNLTLNMGAAHKSQAYRPLLLSTKDGITATYLNDSVD-S 745
Db 697 TTRTSGVGVVMGNQPNFSLDQKVALNMGAAHQAQYRPLWVSTKDGVAITYATDADASKA 756
QY 746 ROYKYTDSOGNLSFSAELQSVANQAQVSGMIQVWPVGAADNQDVRTSPSTQATKDGNIY 805
Db 757 GLVKRTDENGILYFLNDDLKGVANQVSGFLQVWPVGAADQDDIRVAASDTASTDGKSL 816
QY 806 HQSDALDSOVIYEGSFNSFOAQSPOQYNAVIAKNGDLFKSWGITQFEMAPQYVSSBDG 865
Db 817 HQDAAMDSRVMEGFSNFOQSFATKEBEYTNVVIANNVDKFVSWGITDFEMAPQYVSSBDG 876
QY 866 TELDSVILNGYAFSDRYDLAMSKNNKYSKQDLANAICLOSAGIKVLSDLVNPQLYNLP 925
Db 877 QELDSVIQNGYAFSDRYDLAMSKNNKYSKQDLVKAIKALHAKGKLVADWVDPQDYFFP 936
QY 926 GKEVVTATRVNOYQAKSGATINKTPYVANTRSYG-DYOEQYGGKFLDLOLQKLYPRLPST 984
Db 937 KQEVVTVTRDFGKPIAGSQINHSLYYDTKSSGDDYQAKYGGAFDLDELKEKYPELFTK 996
QY 985 KOISTGKPIPSVKITNNKSAKYPNGSNILGRGAKYVLSB--GNKYLNILADGKLFPTVLN 1042
Db 997 KOISTQAIDPSVKIKQNSAKYFNGSNILGRGADYVLSQVSNKYFNVAASDTLFLPSSIL 1056
QY 1043 NTYQPOVSANGFISKNGGIIHYLDKNGQVKNRPFKEISGSWYYPDSGKMATKTKIGND 1102
Db 1057 GKVSSEGRIDG-----KGYIYNSSATGQDVQKASFITEAGNLIFYGKGDMYVGAOTINGA 1112
QY 1103 TYLFPNGKOLKEGVYDGK-KAYYYDNGRTWTKNGFVEFRVDGQKWRYPNGGTIAI 1161
Db 1113 NYFFLENGTALRNTIYTDAGNSHYIYANDGKEYEN-GYQOF---GND-WRYFK-DGNMAV 1166
QY 1162 GLVSLD----- 1167
Db 1167 GLTTVDGNVQYFDKGVQAKDKIIIVTRDGKVRYPDQHNGNAATNTFIADKTGHYYLQKD 1226
QY 1168 -----NRTLYFDAYGVQKQVOTVIN-GKSYTFDADQGL----- 1201
Db 1227 GVAVTGAOTVGRKQLYFEANGQVGRGDFVTSDEGLYFYDVDSGDMWTDFTFIEDKAGNWF 1286
QY 1202 ----- 1201
Db 1287 YLKGDAAVTGAOTIRGQKLYFKANGQVKGDIIVKTDGKIRYIDAKSSEQVFNKTVRAA 1346
QY 1202 -----VQTDNANPAP-----QQG-----AGKKLLGDNQWYRKDG 1231
Db 1347 DGKTYVIGNDGVAVDPVSVVKGQTFKQDASGALRFPYNLKGQLVTGSGWYETANHDWVYIQSG 1406
QY 1232 QLLTGEOTIDGQKVPQDNGVQVKGATDASGLVRFYDRDQHQ----- 1276
Db 1407 KALTGEOTINGOHLYPKEDGHQVKGQLVGTGDKVRYYIDANSQDQAFKNSVTVNGKTYFF 1466
QY 1277 -----VKGWYSTSDDDNNWYVNESQGVLTGLQ 1303
Db 1467 GNDGTAQTAGNPKGQTFKDGSDIRFYSMEGQLVTGSGWYENAGQWLYV-KNGKVLTLGLQ 1525
QY 1304 TIDGQTVFPDDKGIQAKGKAVVDENGNLRYFDADSGNMLRLDRWKNVDGNWYFNRNG 1360
Db 1526 TVGSRQVYFDENGIQAKGKAVRSTDGKIRYFDENSGSMITNQWKFYGYGYFFGNDG 1582

RESULT 11
Q55265_STRSL PRELIMINARY; PRT; 1577 AA.
AC Q55265;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Glucosyltransferase precursor.
GN Name=gtfM;
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
```

```
OC Streptococcus.
OX NCBI_TaxID=1304;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=9512197; PubMed=7822030;
RA Simpson C.L., Giffard P.M., Jacques N.A.;
RT "Streptococcus salivarius ATCC 25975 possesses at least two genes
RL coding for primer-independent glucosyltransferases.";
RL Infect. Immun. 63:609-621(1995).
DR EMBL; U35928; AAC1413.1; -; Genomic_DNA.
DR PIR; T30858; T30858.
DR HSSP; P06653; 1H8G.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR004829; C:surface_antigen.
DR InterPro; IPR002479; C:surface_antigen.
DR InterPro; IPR003318; C:glyco_hydro_70.
DR Pfam; PF01473; C:glyco_hydro_1; 4.
DR ProDom; PD153432; C:surface_antigen; 1.
KW Signal; Transferase.
FT SIGNAL 1 38 Potential.
FT CHAIN 39 1577 glucosyltransferase.
SQ SEQUENCE 1577 AA; 175290 MW; 3EFB898A7D3A7BF3 CRC64;

Query Match 41.9%; Score 3026.5; DB 2; Length 1577;
Best Local Similarity 42.6%; Pred. No. 6.7e-140;
Matches 669; Conservative 215; Mismatches 440; Indels 245; Gaps 42;

QY 1 MEKNRLRYKLRHKQKQWVAIGVTTVTLSPFAGGQVVA-----ADTN-----NNDG----- 44
Db 1 MENKVRPKLRHKVKNWVITIGVTTLSWVALAGSLLAQCKVEADETSAPNGDGLQQLSBDG 60
QY 45 -----TSVQVKNMVPSPKFDQAQNGQLAQMFKAAQNA-----DOTATSVQSPATDGRV- 95
Db 61 TASLVTTTTEQASQAQSAVSASVATSVSHETSFOATSAVSQEATAQTSFVASQEVA 120
QY 96 -----DNOVTP--AANQPAANVANO-----DVANP--ATDAGALNRQSA 131
Db 121 VSSQTQSSGQETQTTEQVSGQSTSTQVAGQTSAQSTPSTEQARPRVLITNAAPATRAA 180
QY 132 D-----TSTDG----- 137
Db 181 DSTIRINARNRNTIITASGITPNTVITITGNTPKPNVTITSPNGRPNVTIVTQPNQN 240
QY 138 KAV-PQTSQDP-----GHLETVDGKTYVDANGQRLAKNYSWIDGKTY 179
Db 241 KPVQPSQPSQPNKPVQPNQPSLDYKPVASNLKTIIDGQYYVE-NGVVKNAAIELDGRLY 299
QY 180 YFDGQTGAQTDLPKTGQANODNVPDS--YQANNQAYSNEASSFFETVDNYLTADSWYRPR 237
Db 300 YFD-ETG-AMVDQSKP-LYRADAI PNNSIYAVYNOAYDTSSKSFHELDNFLTADSWYRPK 356
QY 238 KILKNGQSQWASSEGLRPIILMTWMPDAATKAAAYANFWAKEGLISGSYRQN--SANLDA 295
Db 357 QILKDGKWTASTEKDYRPLMTWMPDKVTQVNYLNYSQQGFGNKTYTTDDMMSYDLAAA 416
QY 236 TQNTQSALEKKITASEGNTNWLDRKMSQFVKSONQWSIASENETVYPNQDHQGGALLFSN 355
Db 417 AETVQRGTEERTGREGNTTLRLQMSDFIKTPQPGWNSESE-DNLLVGDHDLQGGALTFLN 475
QY 356 SKDTEHANSRDLNLRNPTFQTKQY-FTTNYAGYELLNLDVNSPNVQAEOLNHLH 414
Db 476 NSATSHANSDFLNMRTPTNQTGRKYHIDRNSNGYFELLANDIUNSPAVQAEQLNWLH 535
QY 415 YLMWNGDIVMGDKDANFDGVRVDAVDVNNADLLQIQRDYIKAKYGTQDQKNAIDHLSTL 474
Db 536 YIMNIGSILGNDPSANFDGVRIDAVDNVDADLLQIASDYFKYKRVADNEANAIAHLSTL 595
QY 475 EAWSGNDNDYKQDNQNFSLSDINDORSGLMKAFGVASAYRGNLSNLTATAGLNKRSANPDS 534
Db 596 EAWSYNDHQYNKDTKGAQLSDINDPLRETLITTLFLKSNYRGLSERVITNSLNNRSQEQH 655
```



```
QY 659 SYKQDSSGILTSVRVYKKGANSAGNTETRTNSGALLINRPNFR--AYRNLTINMGA 716
Db 698 RNOQVGNSEIITSVRVYKKGALKATDGTDRRTTSGVAVIEGNNPRLRLKASDRVVNMGA 757
QY 717 AHKSOAYRPLLLSTKGIATYLNDSVDSDRQKYKTDSCGNLSFSASELSQSVANAQVSGMI 776
Db 758 AHKQAYRPLLLTTDNGIKAYHSDQEA-AGLVRYTNDRGELIFTAADIKGYANPQVSGYL 816
QY 777 QWVPVGAADNQVRTSPSTQATKGNIIYHQSALDSQVYGGFNSFOAFQSPQYTN 836
Db 817 GVMVPVGAADQVRAASTAPSTDGKSVHQAALDSRVMFEGFSNFQAFATKCEYTNV 876
QY 837 VTAKNGDLKFSKGIITOPENAPQVSSDGTFLDSVILANGYAPSDRYDLAKSNKKYGSQK 896
Db 877 VTAKNVDKFAEAGVTDVFEWAPQVSGVSTGDSFLDSVIQNGYAFDTRDYLGISPKNYKGTAD 936
QY 897 DLANAIKGLQAGIKVLSLVPNLNPKKEVWVATRVNQYGOAKSGATINKTPVANT 956
Db 937 DLVKAIALHSGIKWADWPDQWYALPEKEVWVATRVNKGITGTPVAGSQIKNTLYVVDG 996
QY 957 RYSG-DYQOYQKGLDLDLQKLYPLRFLSTQKIDPSPVKITNWSAKYFNGSNILGR 1015
Db 997 KSSGKQQAQYKGAFLLELOAKYPELFARKQISTGVMPDPSVKIKQWSAKYFNGTNILGR 1056
QY 1016 GAKYVLSLSE--GNKYLNLADGKFLPTVILANTYGOQVPSANGFISKNGGHIYLDKNGQEVK 1073
Db 1057 GAGYVLLKQATNTYFSLVSDNTFLPKSLVNPNGHTSSSVTGLVFDGKGVYVYSTGNOAK 1116
QY 1074 NRPKEISGSWYFDSCKMATGKTIGNDTYLFPMPNGKQKGGVWYDGKKAY-YVDDNGR 1132
Db 1117 NAFISLGNWYFDPNNGYVMTGAQSLNGANYFSLNGIOLRINAIYDNGKVLSYGNDGR 1176
QY 1133 TWTNKGFBFRVDGQDKRYFNGDGTIAIGLSLDNRILYFDAYGYQVKGQTV-TINGKS 1191
Db 1177 RYEN-GYLLF--GQ-QWRYFO-NGIMAVGLRIHGAVQYFDASGFQAKGQFITADGKL 1230
QY 1192 YTFDADQGLVQTDNANPAQOAG-WKLLGDNQWYKXGQGLLTGCEQITIDGQKVPFQDN 1250
Db 1231 RYFDRDSGNQI-----SNRRFVRNSKGEWFLFDHN-----GVAVTGTVTFNGQRLYFKN 1279
QY 1251 GVOVKGGTATDASGLRFRYDRDQGHQVKGWYSTSDDNWVYVNESOVLTGLQTTIDGQTV 1310
Db 1280 GVOAKGEFTRDAGHLRYDPSNGEVNRVFRNSKGEWFLFDHNGIATVTRVNVGQRL 1339
QY 1311 YFDDKGIQAGKAVMDENGLRYFDADSGNMLRDRW-KNVDDGNWYFNRNGLA 1362
Db 1340 YFKNVGVQAKGLITERKRIKYDPSNGEVNRVYRTSSGNWYFNGDGYA 1392
```

RESULT 14

```
Q55263 9STRE
ID Q55263.9STRE PRELIMINARY; PRT; 1590 AA.
AC Q55263;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE GTF-I.
GN Name=Glucosyltransferase;
OS Streptococcus sobrinus;
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=131d;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 33478;
RA Sato S.;
RT "DNA and amino-acid sequences of water-insoluble-glucan synthetase
produced from Streptococcus sobrinus ATCC 33478."
RL Kagoshima Daigaku Shigakubu Kiyo 16:23-29(1996).
RN [2]
RP PROTEIN SEQUENCE.
RX MEDLINE=91224988; PubMed=1827439;
```

```
RA Mooser G., Hefta S.A., Paxton R.J., Shively J.E., Lee T.D.;
RT "Isolation and sequence of an active-site peptide containing a
RT catalytic aspartic acid from two Streptococcus sobrinus alpha-
RT glucosyltransferases.";
RL J. Biol. Chem. 266:8916-8922(1991).
DR EMBL; D63570; BAA09792.1; -; Genomic_DNA.
DR PIR; A39841; A39841.
DR HSP; P06653; 1GVW.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 3.
DR Pfam; PF02324; Glyco_hydro_70; 1.
SQ SEQUENCE 1590 AA; 176058 MW; 9DF7A3F2C6B4FD43 CRC64;
```

Query Match 41.7%; Score 3018; DB 2; Length 1590;
Best Local Similarity 41.6%; Pred. No. 1.8e-139;
Matches 674; Conservative 210; Mismatches 440; Indels 298; Gaps 38;

```
QY 1 MEKNLYKHKVKQWVAIGVTVT-LSFLAGQVVAADTNNNDGTSVQVNMKVPSPDKF 59
Db 1 MEKNVRFKHKVKRWVTLVSASATMLASALGASVASADTDTSDDSNQA--VVTGD--- 55
QY 60 DAQAQNGQLAQAMFKAANOQDQATSPATDGRVDNQVTPAANQPAANVANQDV---- 115
Db 56 --QTTNNQATDQTSIAA-----TATSEQASASTDAATDQ--ASAAEQTGTTASTDTAAQT 106
QY 116 ---ANPA-----TDAGALNRQSAADTSTDGKAVPQTSQPOGHLETVDGKTYV 160
Db 107 TTNAEAKWPTENEOGFTDEMLAEAKVATAESD--SFPSDLAKMSNVKQVDGKYYY 164
QY 161 DANGORLKNYSVIDGKTYVFDGQGTG---EAQTDLPKTGOANQDNVPSYANNOAYSN 216
Db 165 DQDGNVKNFAVSGDKIYYFD-ETGAYKDTSKVDADKSSSAVQNA-TIFAANNRAYST 222
QY 217 EASSPFTVDNYLTADSWYRPRKILKNGOSWQASSEGDLRPIILMTWMPDAATAKAAVFWA 276
Db 223 SAENFEADVNYLTADSWYRPRKILKNGTWESGKDDFRPLJMANWPDTEKENVYNYM- 281
QY 277 KEGLISG-----SVRQNSANLDAATQNTQSAIEKKIASEGNTNWLDRKMSQFVKSONQWS 331
Db 282 --NLVVGIDKTYTAETSQADLTAAAEVQARIEQKITTEQNTKWLREAIASFVKTPQWN 339
QY 332 TASENETVYPNQDHMOGGALLPSNSKD-TEHANSDWRLNRRNPTQTGK-QKYFTTN--- 386
Db 340 GESEK-----PYDDHLQNGALKFDNQSDLTPTDQSNYRLNRRPTPTQTSGLDSRFTYAND 395
QY 387 -YAGYELLANDVDNSNFPVQAEQLNLHLHLYLMNWGDIIVMGDKDANFDGVVRDVAVDNVAD 445
Db 396 PLGGYEFILLANDVDNSNFPVQAEQLNLHLHLYLNFQSIYAKODADANFDSIRVDAVDNVDAD 455
QY 446 LLQIQDIYKAKYGTQNEKNAIDHLSILEAWSGNDNDYVQDQNNFSLISINDQSGMLUK 505
Db 456 LLQISSDYLKAAYGIDKNNKNNHVSIVEAWSNDTPLYLHDDGNLNMNDKFRLSMLW 515
QY 506 AFGYASAYRGNLSNLATAGLKNSANP-DSDDVPNVVFIRAHDSVQTFIAKIIREKLK 564
Db 516 SLAKPLDKRSGNLNPLHNSLVDREYDREVETVPSYSFARAHDSVQDILIRDIKABI-N 574
QY 565 TNADGLTNLTLDLNLKAFDIYNQDMNATDKVYYPNNLPMAYAMWLQNKDVTVRVYGDNY 624
Db 575 PMSFGY-SFTQEEIDQAFKIYNEDEKTKDCKYTHNVPLSYTLTLITNKGSI PRVYGDNF 633
QY 625 TDNGQYMATKTPFYNAIETLLKGRIRKYVAGQAVSYKQDWSSGILTSVRYGKAGSASDA 684
Db 634 TDDQYMANKTVNYDAIESLLKARMKYVSGQAMQNYQIGNGEILTSVRYGKALKQSDK 693
QY 685 GNTETFRSGMALLINRPNFRA-YRNLTINWCAAHKSQAYRPLLLSTKDGITLYLNDSDV 743
Db 694 GDATTTSGVGMVGNQPNFSLDGKGVVALNMGAAHANOEYRALMVSTKDGVTATYATADA 753
QY 744 D-SRQYKYTDSQGNLSFSASELQSVANAQVSGMIQVWVPVGAADNQDVRTSPSTQATKDG 802
```


Db 765 PLLLTDTDKGLQKYLNDSDTNLT--KVADKDGFTPKGSEIKGYKQVEVNGYLSVWVPVGA 822

Qy 785 ADNQVRTSPSTOATKQGN-IYHQSDALDSQVIYEGFSNFQAPQSPDQYTNAVIAKNGD 843

Db 823 KSDQDIRVAASTKANAKGDKSVTASQALDSQLIYEGFSNFQDFVQKDAQYTNKKIAENTD 882

Qy 844 LFKSWGITOFEMAPQVYVSSDGTFLDSVILNGYAFSDRYDLAMSKNNKYGSKQDLANAIK 903

Db 883 LFKANGVTSFEMAPQVVSATDGTFLDSIIQNGYAFSDRYDLAMSKNNKYGSKEDLANALK 942

Qy 904 QLOSAGIKVLSLVPNOLYNLPCKEWTATRYNOYGOAKSGATINKTPYVANTRSYG-DY 962

Db 943 ALHRAGIOAIADWVPDQIYQLPCKEVTASRVNDYGRVKIDQPMWNKLYLANTKSGKDF 1002

Qy 963 QBQYGGKFLDDIQLKLYPLRFLSTKQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKYVLS 1022

Db 1003 QAKYGGFLAEIQLQYPEMFTAKMISTGKPIDSSVKLKWSAQYFNGTNNVLGRGTDYVLS 1062

Qy 1023 -EG-NKYLNLADGKFLPTVLANITYQPOVSANGFISKNGGIHYLDKNGQEVKNRFEKS 1080

Db 1063 DEGTGKYFTVNBKGFLPAVLTG-----DKEAKTGFYNDGKGMTYFTTAGSQAKSDFVTVA 1118

Qy 1081 GSWYFYDSGKWATKTKIGNDTYLFMPNGKOLKEGVWYDGK-KAYYYDDNGRTW----T 1135

Db 1119 GNTYFYDYTGHWVTGPNGINTKPYFPLPNGVMLKDAVMEDDRGRSVYKGTGMYKGSRN 1178

Qy 1136 NKGFEVFRVDGQDKWRYFNGDGTIAIGLVSLDNRTLYFDAYGYQVKGTQVTIN-GKSYTF 1194

Db 1179 NEWFAMTDSKGQLRFRHPDNYGFMSVGLVTIHGNVQYDEEGFQVKGDFTDKAGQTRYF 1238

Qy 1195 DADQGLVQTDNANPAPQCAQKWLKLDNQWGYRKD-GQLLTGQTIIDQKVFPQ-DNGV 1252

Db 1239 DKNTGNLVKGQPFNQ-----NGHWYYSDDQGLIAKGAQTIKQKLYFDAKTGA 1286

Qy 1253 QVKGGTATDASGLRFYDRDQGHQVKGWYSTDDNWVYVNESGOVLTLGLTIDGQTVYF 1312

Db 1287 QVKGDFVTDKGNWTFYSGTGDGLAVSTFFSTGNNAWFYADENGHVAKGEKTIINGQKLYF 1346

Qy 1313 DDK-GIQAKGKAVMDENGNLRYFDADSGNMLRDW---KNVDGNWYFFNRNGLATR 1364

Db 1347 DTKTGOAKGRFVRDAKG-LRFYDADTGALVTNSFLETKAGSNQWYIMGADGYAVR 1401

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 18:56:56 ; Search time 151.097 Seconds
(without alignments)
4414.224 Million cell updates/sec

Title: US-10-797-821-40

Perfect score: 7928

Sequence: 1 MENKHYKLHKVKQWVIA.....PGVYVFDKNGLAYPRVLN 1518

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_21.*
1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001s.*
5: Geneseq2002s.*
6: Geneseq2003as.*
7: Geneseq2003bs.*
8: Geneseq2004s.*
9: Geneseq2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7928	100.0	1518	7	Add93660 Streptoco
2	7928	100.0	1518	9	Adx37283 Streptoco
3	3818	48.2	1554	7	Add93658 Streptoco
4	3818	48.2	1554	9	Adx37281 Streptoco
5	3514.5	44.3	1430	5	Aau98044 S. mutans
6	3510.5	44.3	1430	5	Aau98045 S. mutans
7	3509.5	44.3	1430	5	Aau98041 S. mutans
8	3507.5	44.2	1430	5	Aau98043 S. mutans
9	3505.5	44.2	1430	5	Aau98042 S. mutans
10	3502.5	44.2	1430	5	Aau98029 S. mutans
11	3502.5	44.2	1430	9	Add93656 Streptoco
12	3502.5	44.2	1430	7	Adx37279 Streptoco
13	3325.5	41.9	1590	7	Add93657 Streptoco
14	3325.5	41.9	1590	9	Adx37280 Streptoco
15	3296.5	41.6	1592	2	Aar32925 Glucosylt
16	3126.5	40.3	1475	5	Aau79284 Streptoco
17	3126.5	40.3	1475	5	Aau98027 S. mutans
18	3196.5	40.3	1475	7	Add93654 Streptoco
19	3196.5	40.3	1475	9	Adx37277 Streptoco
20	3195.5	40.3	1475	5	Aau98030 S. mutans
21	3192.5	40.3	1475	5	Aau98040 S. mutans
22	3191.5	40.3	1475	5	Aau98031 S. mutans
23	3190.5	40.2	1475	5	Aau98035 S. mutans
24	3190.5	40.2	1475	5	Aau98033 S. mutans

25	3189.5	40.2	1475	5	AAU98032	S. mutans
26	3188.5	40.2	1475	5	AAU98036	S. mutans
27	3184.5	40.2	1475	5	AAU98034	S. mutans
28	3182.5	40.1	1475	5	AAU98037	S. mutans
29	3172.5	40.0	1475	5	AAU98038	S. mutans
30	3169.5	40.0	1475	5	AAU98039	S. mutans
31	3155	39.8	1577	2	AAU91047	Alpha-D-g
32	3070	38.7	1375	5	AAU98028	S. mutans
33	3070	38.7	1375	5	AAU79288	Streptoco
34	3070	38.7	1375	7	ADU93655	Streptoco
35	3070	38.7	1375	9	ADU937278	Streptoco
36	3062.5	38.6	1499	7	ADU54806	Protein S
37	2883	36.4	2835	5	ABU98574	Dextran s
38	2883	36.4	2835	6	ABU55594	Amino aci
39	2804	35.4	1365	7	ADU93659	Streptoco
40	2804	35.4	1365	9	ADU37282	Streptoco
41	2795.5	35.3	1497	6	ABU63234	Glucanuc
42	2759.5	34.8	1527	5	AAU80055	Leuconost
43	2750.5	34.7	1527	7	ADU54807	Leuconost
44	2601.5	32.8	1477	9	ADU72732	Mutant de
45	2599.5	32.8	1477	9	ADU72696	Mutant de

ALIGNMENTS

RESULT 1
ADD93660
ID ADD93660 standard; protein; 1518 AA.
XX AC ADD93660;
XX AC
DT 29-JAN-2004 (first entry)
XX DT
DE Streptococcus salivarius glucosyltransferase-I.
XX DE
KW Glucosyltransferase; enzyme; vaccine; antitoxin; epitope; immunogen.
XX KW
OS Streptococcus salivarius.
XX OS
PN WO2003075845-A2.
XX PN
PD 18-SEP-2003.
XX PD
PF 07-MAR-2003; 2003WO-US006962.
XX PF
PR 07-MAR-2002; 2002US-0363209P.
XX PR
PR 08-AUG-2002; 2002US-0402483P.
XX PR
(FORS-) FORSYTH INST.
XX
Smith DJ, Taubman MA;
WPI; 2003-845091/78.
XX
Composition useful as vaccines for dental caries comprises a fragment of a glucan binding protein-B binding to a major histocompatibility complex class II protein.
XX
Claim 16; Page 16; 49pp; English.
XX
The present sequence is the protein sequence of Streptococcus salivarius glucosyltransferase-I (GTF-I). Peptide fragments of GTF-I, especially from the catalytic domain of the polypeptide, can be used in immunogenic compositions and subunit vaccines for dental caries. These compositions comprise a major histocompatibility complex (MHC) class II protein-binding peptide from S. mutans glucan binding protein-B (GbpB) covalently linked with a peptide fragment of a streptococcal glucosyltransferase. The compositions are used in a claimed method of eliciting production of an antibody in a mammal. Dieptopic or multiptopic polypeptides can be prepared synthetically or by recombinant DNA technology. Antibodies raised against MHC class II binding fragments of GbpB can be used in passive immunisation.

XX New composition comprising a fragment of a glucan binding protein-B
PT (GbpB) that binds to MHC class II protein, and a biocompatible
PT microparticle, useful for producing an antibody (claimed) for immunizing
PT mammals against dental caries.

XX Claim 7; SEQ ID NO 40; 73pp; English.

XX The invention relates to a composition comprising a fragment of a glucan
CC binding protein-B (GbpB) and a biocompatible microparticle, where the
CC fragment binds to a major histocompatibility complex (MHC) class II
CC protein. The composition is useful for producing an antibody for
CC immunizing mammals against dental caries. This sequence corresponds to a
CC Streptococcus salivarius GbpB protein of the invention.

XX Sequence 1518 AA;

Query Match 100.0%; Score 7928; DB 9; Length 1518;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1518; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MENKIHYKLHKVKKQWTTIIVASVALATVLGGLSVTTSSVSADETQDKTQTQNSGTTAS 60
DB 1 MENKIHYKLHKVKKQWTTIIVASVALATVLGGLSVTTSSVSADETQDKTQTQNSGTTAS 60

QY 61 LVTSPEATKEADKRTNTKEADVLTPAKETNAVETATTTNTQATAEAATATTADVAVA 120
DB 61 LVTSPEATKEADKRTNTKEADVLTPAKETNAVETATTTNTQATAEAATATTADVAVA 120

QY 121 PNKEAVTTDAPVTTTEKAEQPAATVKAEEVNTVEVKAPEALAKDSEVEAALSUKNTK 180
DB 121 PNKEAVTTDAPVTTTEKAEQPAATVKAEEVNTVEVKAPEALAKDSEVEAALSUKNTK 180

QY 181 GKYYVNEGSHKENFAITVNGQLYFGKDGALTSSTSYSTFTPTNTNIVDGFSSINRAYD 240
DB 181 GKYYVNEGSHKENFAITVNGQLYFGKDGALTSSTSYSTFTPTNTNIVDGFSSINRAYD 240

QY 241 SSEASPELIDGYLTADSWRPAASIIKDGVTWQASTAEDFRPLMAMWPNVDQVNYLNM 300
DB 241 SSEASPELIDGYLTADSWRPAASIIKDGVTWQASTAEDFRPLMAMWPNVDQVNYLNM 300

QY 301 SKVFNLDAKYSSTDKQETLKVAAKDIQIKIEQIKAEKSTQWLRETIISAFVKTQPOW 360
DB 301 SKVFNLDAKYSSTDKQETLKVAAKDIQIKIEQIKAEKSTQWLRETIISAFVKTQPOW 360

QY 361 TENYSKGGEDHLQGGALLYVNDSTRTPWANSYRRRLNRTATNQTGTIDKSIILDEQSD 420
DB 361 TENYSKGGEDHLQGGALLYVNDSTRTPWANSYRRRLNRTATNQTGTIDKSIILDEQSD 420

QY 421 MGGFDFLLANDVLSNPVQAEQLNQIHYLMNWSIVMGDKDANFDGIRVDVNDVADM 480
DB 421 MGGFDFLLANDVLSNPVQAEQLNQIHYLMNWSIVMGDKDANFDGIRVDVNDVADM 480

QY 481 LQLYTNYFREYGVNKSEANALAHISVLEAWSLNDNHNDKTDGAALAMENKQRLALLFS 540
DB 481 LQLYTNYFREYGVNKSEANALAHISVLEAWSLNDNHNDKTDGAALAMENKQRLALLFS 540

QY 541 LAKPIKERTPAVSPLYNNFTNTORDEKTDWINKDGSKAYNEDGTVKQSTIGKYNEKY 600
DB 541 LAKPIKERTPAVSPLYNNFTNTORDEKTDWINKDGSKAYNEDGTVKQSTIGKYNEKY 600

QY 601 ASGNTVFIIRAHNNVDIIAEIIKKEINPKSDGFTTDAEMKQAPRIYNKMLSSDKK 660
DB 601 ASGNTVFIIRAHNNVDIIAEIIKKEINPKSDGFTTDAEMKQAPRIYNKMLSSDKK 660

QY 661 LNNIPAAVAMLQNMETITRVYVGDLYTDGHHMETKSPYYDTIVNLMKSRIKYVSGQA 720
DB 661 LNNIPAAVAMLQNMETITRVYVGDLYTDGHHMETKSPYYDTIVNLMKSRIKYVSGQA 720

QY 721 QRSYWLPTDGRKMDNSVELYRNEVYTSVRYGKDMTANDTEGSKYSRTSGQVTLVANNP 780
DB 721 QRSYWLPTDGRKMDNSVELYRNEVYTSVRYGKDMTANDTEGSKYSRTSGQVTLVANNP 780

QY 781 KLNLDQSAKLNVMEMGIHANQKRYALIVGTADGIKNFTSDADAIAAGYVKETDSNGVLTF 840
DB 781 KLNLDQSAKLNVMEMGIHANQKRYALIVGTADGIKNFTSDADAIAAGYVKETDSNGVLTF 840

QY 841 GANDIKGYETFDMSGFVAVWVPVGASDNQDIRVAPSTEAKKEGELTLKATEAYDSQLIYE 900
DB 841 GANDIKGYETFDMSGFVAVWVPVGASDNQDIRVAPSTEAKKEGELTLKATEAYDSQLIYE 900

QY 901 GFSNFOTIIPDGSPPSYVTNRKIAENVDLFKSMGWTSFEMAPQVVSADGDTFLDSVQNGY 960
DB 901 GFSNFOTIIPDGSPPSYVTNRKIAENVDLFKSMGWTSFEMAPQVVSADGDTFLDSVQNGY 960

QY 961 AFADRYDLAMSKNKNYKSKEDLRDALKALHKAGIQAIADWPDQIYQLPGKEVVTATRTD 1020
DB 961 AFADRYDLAMSKNKNYKSKEDLRDALKALHKAGIQAIADWPDQIYQLPGKEVVTATRTD 1020

QY 1021 GAGRKTIADAIIDHSLYLVANSKSGKDYQAKYGGEFPLAELKAKYPENFKVNMISTGKPIDD 1080
DB 1021 GAGRKTIADAIIDHSLYLVANSKSGKDYQAKYGGEFPLAELKAKYPENFKVNMISTGKPIDD 1080

QY 1081 SVKLLKQWKAIEYFNGTIVLERGVYVLSDEATGKYFTVTKEGNFIPLQLTGKEKVIITGFSS 1140
DB 1081 SVKLLKQWKAIEYFNGTIVLERGVYVLSDEATGKYFTVTKEGNFIPLQLTGKEKVIITGFSS 1140

QY 1141 DGKGIITYFCTSGTQAKSAFVTFNGNTIYFDARGHMVTNSEYSPNGKDVYRFLPENGIMLSN 1200
DB 1141 DGKGIITYFCTSGTQAKSAFVTFNGNTIYFDARGHMVTNSEYSPNGKDVYRFLPENGIMLSN 1200

QY 1201 APYIDANGNTIYLYNSKGQMYKGYTKFDVSETDKGKESKVVKPRFTNEGVNAKGVTVI 1260
DB 1201 APYIDANGNTIYLYNSKGQMYKGYTKFDVSETDKGKESKVVKPRFTNEGVNAKGVTVI 1260

QY 1261 DQFTQYFGBDGFQAKDKLVTFKGTYYFDAHTGNGIKDWTNRNINGKWIYFDANGVAATGA 1320
DB 1261 DQFTQYFGBDGFQAKDKLVTFKGTYYFDAHTGNGIKDWTNRNINGKWIYFDANGVAATGA 1320

QY 1321 QVINGQKLYFNEBDSQVKGKGVKNADGTYSKYKEGFGELVTNEFTTDTGDNVWYIYAGANGK 1380
DB 1321 QVINGQKLYFNEBDSQVKGKGVKNADGTYSKYKEGFGELVTNEFTTDTGDNVWYIYAGANGK 1380

QY 1381 TVTGAQVINGQHLIYFNADSGVKGKGVKNADGTYSKYNAKSTGERLTFNEFTTGDNNWYI 1440
DB 1381 TVTGAQVINGQHLIYFNADSGVKGKGVKNADGTYSKYNAKSTGERLTFNEFTTGDNNWYI 1440

QY 1441 GANGKSVTGEVKIGDDTFFPAKDGKQVKGQTVSAGNGRISYYGDSGKRAVSTWIEIQPG 1500
DB 1441 GANGKSVTGEVKIGDDTFFPAKDGKQVKGQTVSAGNGRISYYGDSGKRAVSTWIEIQPG 1500

QY 1501 VVYVFDKNGLAYPPRVLN 1518
DB 1501 VVYVFDKNGLAYPPRVLN 1518

RESULT 3
ADD93658
ID ADD93658 standard; protein; 1554 AA.
XX
AC ADD93658;
XX
XX 29-JAN-2004 (first entry)
XX
DE Streptococcus sobrinus glucosyltransferase-U.
XX
KW Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.
XX
OS Streptococcus sobrinus.
XX
PN WO2003075845-A2.
XX
PD 18-SEP-2003.
XX
XX 07-MAR-2003; 2003WO-US006962.
XX

XX AC ADX37281;
 XX DT 21-APR-2005 (first entry)
 XX DE Streptococcus sobrinus glucan binding protein B #2.
 XX KW immunogenicity; immune stimulation; glucan binding protein-B;
 XX OS microparticle; major histocompatibility complex; tooth disease.
 XX OS Streptococcus sobrinus.
 XX PN US2005031633-A1.
 XX PD 10-FEB-2005.
 XX PF 09-MAR-2004; 2004US-00797821.
 XX PR 13-APR-1998; 98US-0081550P.
 XX PR 08-JAN-1999; 99US-0115142P.
 XX PR 12-APR-1999; 99US-00290049.
 XX PR 07-MAR-2002; 2002US-0363209P.
 XX PR 08-AUG-2002; 2002US-0402483P.
 XX PR 07-MAR-2003; 2003US-00383930.
 XX (SMIT/) SMITH D J.
 XX (TAUB/) TAUBMAN M A.
 XX Smith DJ, Taubman MA;
 XX WPI; 2005-151644/16.
 XX New composition comprising a fragment of a glucan binding protein-B
 XX (GbpB) that binds to MHC class II protein, and a biocompatible
 XX microparticle, useful for producing an antibody (claimed) for immunizing
 XX mammals against dental caries.
 XX Claim 7; SEQ ID NO 38; 73pp; English.
 XX The invention relates to a composition comprising a fragment of a glucan
 XX binding protein-B (GbpB) and a biocompatible microparticle, where the
 XX fragment binds to a major histocompatibility complex (MHC) class II
 XX protein. The composition is useful for producing an antibody for
 XX immunizing mammals against dental caries. This sequence corresponds to a
 XX Streptococcus sobrinus GbpB protein of the invention.
 XX Sequence 1554 AA;
 XX Query Match 48.2%; Score 3818; DB 9; Length 1554;
 XX Best Local Similarity 50.1%; Pred. No. 3.9e-196;
 XX Matches 783; Conservative 215; Mismatches 434; Indels 130; Gaps 24;
 QY 1 MENKTHYKLUKVKQWVTTIIVASVALATVILGSLVTSSTSSVADETQDKVTIVTOSNGTTRAS 60
 DB 1 MEKGLUHLKLUKGVKHWVTTIIVASIGLVSLV-----AGTVSA---EDKVANDTTAQATVG 52
 QY 61 LVTSPEATKEADKRTNKADVLTPAKETNAVETAT-TTNTQATAEAAATTATTADVA--VA 118
 DB 53 VDTGQDQATNTDANTNTDTTADQASANTNQAGSDQSNQDQAKQDTANTDRNQADNS 112
 QY 119 AVPNKEAVVTTDAPAVTTTEKAEQEPATVKAEEVNTTEVKAPEALKDSEVEAALSLEKNIKN 178
 DB 113 QTDNNQATDQATSPATDGTSGVQRDA--ANVATAADQSGQTAPSEQEKSAALSLEKNIKN 169
 QY 179 IDGKYVYNNEDSHKENFAITVNGQLLYFGKD-GALTSSSTTSFPTGTTNIVDGFSINNR 237
 DB 170 IDGKYVYVQADSGSYKKNFAITVNGQMLYFSDTSFGALSTSTYSFSGGTTNLVDVDFSSHNK 229
 QY 238 AYDSSEASPELLDGLVTADSWRPSAIIKDGVTWQASTAEDRPILLMAWPNVDTQVNYL 297
 DB 230 AYDSTAKSFELVNGVLTANSWRPAGILRNQGTWEASNENDURPVLMSWMPDKOTQVAY 289
 QY 298 NYMSKVFNL-DAKYSSTDQKQETLKVAAKDIIQIKIEQIAQKSTQWLRETTISAFVKTQPO 356

DB 290 NYMKYLSANETEVNTNETSQVDLNLKEAQSIQTKIEQITSDNSTQWLRTAMEAFVAAQPK 349
 QY 357 WNKETENYSGGGEDHLQGGALLYYND SRTPWANSDYRLNRATNTOTGTIDKSILDEQS 416
 DB 350 WNMSTENFKG---DHLQGGALLYYND SRTPWANSDYRLNRATNTOTGTIDKSILDEQS 404
 QY 417 DPNHMGGFDFLLANDVDLSNPVQAEQLNQIHYLMNMGSIVMGDKDANFDGIRVDADVNV 476
 DB 405 E---GGYEFLLSNDVNSNPVQAEQLNQIHYLMNMGDIVMGDKDANFDGIRVDADVNV 460
 QY 477 DADMLQLYNYFREYGVNKSSEANALAHISVLEBMSLNDNHNKTDGGAALAMENKQRLA 536
 DB 461 NADLLQVSNYFKDNYKVTDSEANALAHISILEAWSLNDNQYNEDTNGTALSIDNSSRLT 520
 QY 537 LLFSLAKPIKERTPAVSPLYNNTTORTDEKTDINKDGSKAYNEDEGTQVKQSTICKYNE 596
 DB 521 SLAVLTK-----QPGQRIDISNLISESVNKERAND-----T 551
 QY 597 KYGDASGNVYFIRAHNNVQDIIAEIIEKEINPKSDGFTITDAEMKQAFIYNKMDLSSD 656
 DB 552 AYGDITPTYSFVRAHDSEVQTVIAKIVKEIDTNSDGYFTFLDQLKDAFKIYNEDMAKVN 611
 QY 657 KKYTLNNIPAAVAVMLQNMETITRVYVYDLYTDDGHYMETKSPYYDTIYNLMKSKRIYVS 716
 DB 612 KTYTHYNIPAAVALLSNMESVPRVYVYDLYTDDGQYMAKKSPPYDAIATMLQGRILAYVS 671
 QY 717 GGOAQRSYWLPTDGRKMDNSDELVELYRTNEVYTVSRVYCKDITWANDTEGSKYSRTSQVTLV 776
 DB 672 GGOSEVH-----KVGNNQILSSRYGQDLMSADDTQGTDLSTSGVLTLV 718
 QY 777 ANPNKLNLDQSAKLVNEMKIHANOKYRALIVGTADGIGNFTSDADAIAAGVYKETSNSG 836
 DB 719 SNDPNLDLGGDS-LTVNMGRAHANQAYRPLILGTGQVQSYLKSDS--TNIVKYTDANG 774
 QY 837 VLTFGANDIKGYETFDMSGFVAWVVPVGASDNQDIRVAPSTBAKKEGELTLKATEAYDSQ 896
 DB 775 NLTFADDIKGYSTVDMGYLAVVPVGAQGVVAAQDVRVAADTNQKADGK-SLAKTSAALDSQ 833
 QY 897 LIYEGFSNFQITPDGSDPSVYTNKIAENVDLFKSWGTSFEMAPOFVSADGTFPLDSVI 956
 DB 834 VIYEGFSNFQDF--ANNADYTNKXIAENADFFKLGITSFEMAPOFVSATDGSFLDSII 891
 QY 957 QNGYAFADRYDLAMSKNNKYGSKEDLRDALKALHKAGIOAIADWVPDQIYQLPGKEVWTA 1016
 DB 892 QNGYAFADRYDLAMSKNNKYGSKEDLRDALKALHKAGIOAIADWVPDQIYQLPGKEVWTA 951
 QY 1017 TRTDGAGRKIADAIIDHSLYVANSKSGKYOQAKYGGEGFLABELKAKYPMFKVMNISTGK 1076
 DB 952 KRTNSYGNPTFDAYINNALYATNTKSSGSDYQAQYGGAFDELKAKYPMFTVMNISTGK 1011
 QY 1077 PIDDSVKLKWKAERYFNGTNVLERGVYVLSDEATKGYFTVTKEGNFIPLOLTGKEKVT 1136
 DB 1012 PIDPSTKIKWEAKYFNGTNVLERGVYVLSDEATKGYFTVTKEGNFIPLOLTGKEKVT 1071
 QY 1137 QPSSDGKIGTYGTSQAKSAFVTFNGNTYVFDARGHVNTNSEYSPNGKDVYRPLPNCI 1196
 DB 1072 GFYDGTGMAYTSGNKAVNSFIYEGGHYFFDQGHVWTSYKAEADGNDYY-FLPNCI 1130
 QY 1197 MLSNIFYIDANGNTYLYNSKQOMYKGG--YTKFDVSETDKGCKSVKVKFRYFTNEGVA 1254
 DB 1131 QMRDAIYQDAQNSYYGRTGILYKGNWPPVDPNNANK-----TVFRYFDANNVMA 1183
 QY 1255 KGVTVIDGFTQYFGEDEGQAKDLVTFKGYTYFPAHTGNGIKDTRNINGWYVYFDANG 1314
 DB 1184 IGYRMYGQTYFDENGFQAKQLLTDKGTHTYFDEDNAGAMAKNFVNVGDDMYMDGNG 1243
 QY 1315 VAATCAQVINGQKLYFN-EDGSQVKGGVVKNADGTSYKKEGFLVNTERTTGNVWY 1373
 DB 1244 NAVKQYVNNQILYFNPTGVQVKGQFITDQGRTSYDANSALKSGGFFTPGSDWY 1303
 QY 1374 YA-----GANGKTVTGAQVINGQHLFYFNAD-GSQVKGGVKN 1409

Db 1304 YAENGVVYKGFQVAENQDQWYYFDQTTCKQAKGAQKRDLDYFNPDSGVQVKGDFATD 1363

Qy 1410 ADGTVSKYNASTGERLTNEFFTTGDNWYIYGANGKSVTGEVKI----- 1453

Db 1364 ESGNTSFYHGDNGDKVVGGFTTGNNAWYADNNGNLVKGFEIDGKWHFDEVTGQQA 1423

Qy 1454 -----GDDTYFAKDGQVKGQTVSAGNRSYYYGDSGRKAVSTWIEIQGVVYVDPK 1507

Db 1424 GAALVNGQQLYFDPVDSIGIQVKGDFVTDGQGNYSYYDVNSGDKKVGFFTTGDNAMYADG 1483

Qy 1508 NG 1509

Db 1484 QG 1485

RESULT 5

AAU98044

ID AAU98044 standard; protein; 1430 AA.

XX AAU98044;

AC AAU98044;

XX 27-AUG-2002 (first entry)

XX S. mutans glucosyltransferase GTFD mutant N471D/T589D.

DE Glucosyltransferase; GTFD; transgenic plant; paper sizing;

XX coating composition; glucan; starch; latex; thermoplastic molecule;

KW amyloplast; vacuole; paper manufacture; mutant; muten.

XX Streptococcus mutans.

OS Synthetic.

XX

XX Key Location/Qualifiers

FH Misc-difference 471

FT /note= "wild-type Asn substituted by Asp"

FT Misc-difference 589

FT /note= "wild-type Thr substituted by Asp"

XX US2002031826-A1.

PN 14-MAR-2002.

XX

PD 19-DEC-2000; 2000US-00740274.

PF

XX 07-JUN-1995; 95US-00478704.

PR 07-JUN-1995; 95US-00482711.

PR 07-JUN-1995; 95US-00485243.

PR 16-JAN-1998; 98US-00007999.

PR 16-JAN-1998; 98US-00008172.

PR 20-JAN-1998; 98US-00009620.

PR 11-DEC-1998; 98US-00210361.

XX (NICH/) NICHOLS S E.

PA Nichols SE;

PI

XX WPI; 2002-414332/44.

DR

XX Glucosyltransferase B or D protein useful for producing a glucan useful

PT as substitutes for and additions to modified starch and latexes in paper

PT manufacture, comprises mutations in specific positions.

XX

PS Claim 36; Page: 44pp; English.

XX The invention an isolated protein comprising a glucosyltransferase (GTF)

CC B polypeptide having changes at position from I448V, D457N, D567T,

CC K1044T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,

CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a

CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,

CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the

CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its

CC complementary polynucleotide, a ribonucleic acid sequence encoding the

CC GTF mutant, an expression cassette comprising the polynucleotide operably

CC linked to a promoter, a vector comprising the expression cassette, host

CC cell introduced with the vector, a transgenic plant comprising the

CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or

CC coating composition comprising a glucan produced in a plant transformed

CC with a gene encoding the mutant GTF, wild type or, starch, a latex,

CC thermoplastic molecule or their combinations or glucan and starch where

CC the glucan is produced in the amyloplast and/or vacuole or a maize line

CC deficient in starch biosynthesis, transformed with a gene encoding a

CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper

CC comprising the glucan (paper sizing/coating agent). The vector is useful

CC for producing a glucan in a plant. The method comprises transforming a

CC plant cell with the vector, growing the plant cell under plant growing

CC conditions to produce a regenerated plant and inducing expression of the

CC polynucleotide for a time sufficient to produce the glucan in the

CC regenerated plant, where the vector contains a transit sequence from

CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and

CC chlorophyll AB binding protein to produce a transgenic plant, and glucan

CC is produced in the amyloplast of potato or the vacuole of sugar beet.

CC Glucans are useful as substitutes for and additions to modified starch

CC and latexes in paper manufacture. Unlike prior art techniques, which

CC require input materials that produce chemical effluents, paper

CC manufacture utilising the glucan produced by GTF, which utilises

CC biologically produced input materials, is more cost-effective and

CC environmentally friendly. Moreover, glucans also exhibit thermoplastic

CC properties and impart gloss to the paper during coating step. The present

CC sequence represents a GTF mutant of the invention. Note: The present

CC sequence is not shown in the specification but was created by the Indexer

CC using the GTFD sequence appearing as AAU98029 and the information in

CC claim 36

XX

SQ Sequence 1430 AA;

Query Match 44.3%; Score 3514.5; DB 5; Length 1430;

Best Local Similarity 48.2%; Pred. No. 7.1e-180;

Matches 726; Conservative 246; Mismatches 433; Indels 101; Gaps 32;

Qy 1 MENKHYKLVKVKQWVTVIAVASVALATVLGSLVTTSSVSAD---ETQDKTVTO-SNSG 56

Db 1 METKRYKMKHKVHWVTVAVAS-GLITL--GTTILGSSVSAETQQTSKRVYTKQSEDD 57

Qy 57 TTASLVTSPEATKADKRTNTKEADVLTPAKETNAVETATTTNTQATAEAAAT-ATTADV 115

Db 58 KAASESSQTDAPKTKQAQTEQTAQ-----SQANVADTSTSTTKETPSQNTTQANSDDK 112

Qy 116 AVAAVPNKEAVVTDAPAVTTEKAEQAPATVKAEVVNTVEVKA-----PEALKDSEVTA 169

Db 113 TVTNTKSEEAQTSEE---RTKQSEEAQTASSQAL-TQAKAELTKQRTAAAGKNPNVD 167

Qy 170 ALSLKNIKNIDKYYVNVNEDGSHKENFALTVNGQLLYFGKD-GALTSSTSYFTPTTNI 228

Db 168 LAAPVNVKQIDGKYIYIGSDGQPKNFALTVNNKVLVFDKNTGALTDTDSQYQKGLTKL 227

Qy 229 VDFGSINNRAYDSSEASFELIDGYLTADSWYRPASIIKDGVTWQASTAEDEFPLMAWMP 288

Db 228 NNDYTPHQIVNFENTSLETIDNYVTADSWYRPKDLKNGKTWTASSESDLPRLMSWMP 287

Qy 289 NVDTQVNYLYNYSKV-FNLDAKYSTDKQETLUKAAKQIQIKIEOKIAQEKSTQWLRRTI 347

Db 288 DKQTQIAYLYNMQOGLGTGENYTADSSQESLNLAATVQVKIETKISQTSQOTQWLRLDI 347

Qy 348 SAFVKTOPQWNETENYSKGGEDHLOGGALLYVNDSTRTPWANSDYRLNRTATQGTGI 407

Db 348 NSFVKTPQWNWSQTESDSAGEKDHLOGGALLYSNSDKTAYANSDYRLNRTPTSGTK- 406

Qy 408 DKSILDEQSDPNHMGDFPFLANDVDLSNPVVQAEQLNQIHYLMNWGSIYVMGDKQANFDG 467

Db 407 -----PKYFEDNSSGCGYDFLLANDIDNSNPVVQAEQLNLHYLMNYSIVANDPEANFDG 461

Qy 468 IRVDAVDNVDADMLQYNYFREYGVNKSSEANALAHISVLBAWSLNDNHNDKTDGAAL 527

Db 462 VRVDAVDNVDADLLQIASDYLKAHYGVDKSEKNAINHLSILEAWSNDNDPQYNKDTKGAQ 521

Qy 528 AMENKQRLALLFLSLAKPIK-----ERTPAVPSLYNNTFTNTQORDEKTDWINKDGSKAY 580

CC require input materials that produce chemical effluents, paper
 CC manufacture utilising the glucan produced by GTF, which utilises
 CC biologically produced input materials, is more cost-effective and
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
 CC properties and impart gloss to the paper during coating step. The present
 CC sequence represents a GTFD mutant of the invention. Note: The present
 CC sequence is not shown in the specification but was created by the indexer
 CC using the GTFD sequence appearing as AAU98029 and the information in
 CC claim 36
 XX
 SQ Sequence 1430 AA;

Query Match 44.3%; Score 3510.5; DB 5; Length 1430;
 Best Local Similarity 48.1%; Pred. No. 1.2e-179; Mismatches 433; Indels 101; Gaps 32;
 Matches 725; Conservative 247;

QY 1 MENKHYKLVKVKQWVTVIIVASVALATVGLSLVTSVSAD---ETQDKVTQ-SNSG 56
 DB 1 METKRYKMHVKHVVAVAS-GLITL--GTTILGSSVSAETEQTSQDKVVTQKSEDD 57

QY 57 TTASLVTSPEATKEADKRTNTEADVLTPAKETNAVEATTTNQTATBAAT-ATTADV 115
 DB 58 KAASESSQTDAPKTKQAQTEQTAQ-----SQANVADTSTSTKETPSONITTOANSDDK 112

QY 116 AVAAVPNKEAVVTDAPAVTTEKAEQAPATVKAEEVNVTEVKA-----PEALAKDSEVEA 169
 DB 113 TVTNTKSEEAQTSSE-----RTQSEEAQTTASSQAL--TQAKAELTKQRTQAQENKNPVD 167

QY 170 ALSLNKNIKIDGKYVYVNDGSHKENFALTIVGQLLYFGKD--GALTSSSTYSFTPTNNI 228
 DB 168 LAAPNVKIDGKYVYIGSDGPKKNFALTIVNKVLYFDKNTGALTDTDSQYQPKQLTKL 227

QY 229 VDGFSINNRAYSSEASFLIDGYLTADSWYRPASIIKDGVTWQASTABDFPLLMAMP 288
 DB 228 NNDYTPHNGIVNFENTSLTIDNYVTADSWYRPKDIKNGKTTWASSSEDLPLLMAMP 287

QY 289 NVDTQVNYLYNYSKV--FNLDKAYSSTDQETLKVAAKDIOIKIEQIKQAEKSTOWLRETI 347
 DB 288 DKQQTAYLYNMQOGLGTGENYATDSSQESLNLAQTVQVKIETKISQTOQTOWLRDII 347

QY 348 SAFVKTQPOWNETENYKSGGEDHLOGGALLYVNDSTRPWNSDYRRLNRTATNQTGTI 407
 DB 348 NSFVKTQPNWNSQTESDTSAGEKDLHOGGALLYSNSDKTAYANSYRLLNRTPTSQTK- 406

QY 408 DKSILDEQSDPNHMGDFLLANDVLSNPVQAEOLNOIHYLMNWSIVMGDKDANFDG 467
 DB 407 -----PKYFEDNSSGGYDFLLANDIDNSPNPVQAEOLNWLHYLMNYGSIVANDPEANFDG 461

QY 468 IRVDVNDVADMQLQLYTNYFREYYGVNKSSEANALAHISVLEAWSLNDNHNDKTDGAAL 527
 DB 462 VRVDVNDVNDALLQIASDYLKAHYGVDKSEKNAINHLSILEAWSNDPQYKNDTKGAOL 521

QY 528 AMENKQRLALLFLSLAKPIK-----ERTPAVSPLYNNTFTTQRDEKTDWINKDGSKAY 580
 DB 522 PIDNKLRLSLLYALTFPLEKADSKNEIRSGLEPVITNSLN-----562

QY 581 NEDGTQKQTIKYNKKGADSGNVYPIRAHNNVDIIAEIIKKEINPKSGDFTITDAE 640
 DB 563 -----NRSAGKNSER-----MANYIFIRAHDSVEQVEIAKIIKAQINPKTDGLTFLDE 612

QY 641 MKQAFEPYKNDMLSSDKKYTLNPIPAAYAVMLQNMETITRVYGDLYTDGHHMETKSPY 700
 DB 613 LKQAFKIYNEDMRQAKKYTKQNIPTAYALMISNKDSTIRLYYGDGMSDDGQYMATKSPY 672

QY 701 YDTIVNLMKSRIKYSGGQARSYMLPTDGKMDNSDVELYRTNEVVTYSVRYGKDIMTAND 760
 DB 673 YDAIDTLLKARIKYAAGGDMKITVY--EGDKSHMD---WDYTGVLTSVRYGTGANEATD 727

QY 761 TSGSKYSRTSGQVTLVANNPKLNDQSAKLNVEMGKIHANQYKRALIVGTAGIKNFTSD 820
 DB 728 -QGSATKTQGMAVITSNPNPSLKLNDQKIVIVNMGAHKNQBYRPLLLATTKDGLTSYTS- 785

QY 821 ADAIAAGYVKETDSNGVLTFGANDIKGYETFDMSGFPVAVVVPVGSADNQDIRVAPSTEAK 880

DB 796 -DAAAKSLYRKTNDKGELVFDASDIQGYLNPVQSYGLAVWVPVGSADNQDVRVAASNKAN 844
 QY 881 KEGELTLKATEAYDSQLIYEGFSNFQT-IPDGSDPVVTNRKIAENVDLFSKSGVTSFPM 939
 DB 845 ATGQV-YESSSALDSQLIYEGFSNFQDFVTKDSD--YTNKKIAQNVLFSKSGVTSFPM 900

QY 940 APQFVSADDTFLDSVIONGVAFADRYDLAMSNNKYSKEDRLDALKALHAGIQAAD 999
 DB 901 APQYVSSDGSFLDSIIQNGYAFEDRYDLAMSNNKYSQDDMINAVKALHKSIGIQAAD 960

QY 1000 WYPDOIYOLPGKEVVTATRTDAGRKIADAIIDHSLYVANSKSGKDYQAKYGGGFSLAE 1059
 DB 961 WYPDOIYNLPGKEVVTATRVNDYGEYRKDSKELKNTLYAANTKSNKDKQAKYGGAFSL 1020

QY 1060 KAKYEMFKVNMISTGPKPIDDSVKLQWKAEBYFNGTNNLVRGVYVLSDEATGKFTVTK 1119
 DB 1021 AAKYPSIFNRTQISNGKKIDPSEKITAMKAKYFNGTNNLVRGVYVLSKDNASDKYFELKG 1080

QY 1120 EGNFTPLQLTGKEKVIITGSSDGKGIITFYGTSGTOAKSAFV-TFNGNTYTFDARGHVTN 1178
 DB 1081 NQTYLPKQMTNKE-ASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYTFDNGHMYG 1139

QY 1179 SEYSPNGKDVYRFLPNGLMISNAPYIDANGNTYLYNSKGMKYGGYTKFDVSETDKDGKE 1238
 DB 1140 LQ-QLNG-EVQYFLSNGVQLRESFLENADGSKYFGLHGNRYSNGYYSFDNDS----- 1190

QY 1239 SKVKRYFTNEGVMNAKVTVIDGTYQYFGEDEFOAKDKLVT-FKGKTYTFDAHTGNGIK 1297
 DB 1191 ---KWRYFDASGVMAVGLKTINGNTQYFDQDGYQVKGAWITGSDGKKRYFDGSGNMAV 1246

QY 1298 DTWRN-INGKWIYFDANGVAATGAQVINGQKLYFNEGDSQVKGVVKNADGTYSKYKEGF 1356
 DB 1247 NRPANDKKGDMYLYNSDGLVGVQTINGKTYTFQDQKQIKGKIITD-NGKLKYFLANS 1305

QY 1357 GELVTNEFTTDDGNVWYVAGKVTGTAQVINGOHLVFNADGSOVKGVVKNADGTYSK 1416
 DB 1306 GELARNIFATDSQNNWYTFGSDGVAVTGQTAGKLLYPASDGKQVKSFV-TYNGKVHY 1364

QY 1417 YNASTGERLTNEFTTGDNNWYIIGANGKSVTGEVKIGDDTYFFAKDGKQVKQTVSAGN 1476
 DB 1365 YHADSGELQVNRFEADKGNWYLYLDSNGEALTGSRINDQRVFTTREGKQVKGD-VAYDE 1423

QY 1477 GRISY 1482
 DB 1424 RLLVY 1429

RESULT 7
 AAU98041
 ID AAU98041 standard; protein; 1430 AA.
 XX
 AC AAU98041;
 XX
 DT 27-AUG-2002 (first entry)
 XX
 DE S. mutans glucosyltransferase GTFD mutant TS89D.
 KW Glucosyltransferase; GTFD; transgenic plant; paper sizing;
 KW coating composition; glucan; starch; latex; thermoplastic molecule;
 KW amyloplast; vacuole; paper manufacture; mutant; mutein.
 XX Streptococcus mutans.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 589
 FT /note= "wild-type Thr substituted by Asp"
 XX
 XX US2002031826-A1.
 XX 14-MAR-2002.
 XX


```
QY 587 KQSTIGKYNEKYGDASGNVYFIPRAHNNVODIIAEIIKKEINPKSGFTITDAEMKQAFE 646
Db 549 P-----SYSFARHPDSEVQDIIRDKAEINPNFSGYFTQIEIQAFK 592
QY 647 IYNKDWLSSDKKYTLNIPAAVAVMLQNMETITRVVYGDLYTDDGHYMETKSPYYTIVN 706
Db 593 IYNEDLUKDKDKYTHYVPLSYLLTNKSGIPRVYVYGMFTDDGQYMANKTNYDAIES 652
QY 707 LKMSRIKYVGGQQAORSYMLPTDCKMDNSVELYRTNEVTVSVRYGKDINTANDTSGSKY 766
Db 653 LLKARMKYVGGQAMQNYQI-----GNGEILTSVRYGKALKQSD-KGDAT 697
QY 767 SRTSGQVTLVANNPKLNDQSAKLNVEMKIHANQYRALIVGTADGINKNFTSDADAIAA 826
Db 698 TRTSGVGVVMGNQPNFLDGKV-VALNMGAHAHQBYRALVSTKDGVAITYATDADASKA 756
QY 827 GYVKETDSNGVLTFGANDIKGYETFDMSGFVAWVPVGSADNQDIIVAPSTEAKKEGELT 886
Db 757 GLVKRTDENGYLFLNDDLLKGVANPQVSGFLQWVPVGAADQDIIRVAASDTASTDGK-S 815
QY 887 LKATEYDSQLIYEGFSNFQTTIPDGDSPSVYTNRKIAENVDLFKSMGVTSFEMAPQFVSA 946
Db 816 LHQDAAMDSEVMFEGFSNFQSF--ATKEBEYTNVIANVNDKPVSVNGIITDFEMAPQYVS 873
QY 947 DGTFTLDSVIQNGYAFADRYDLAMSNNKYGSKEDLRDALKALHKAGIQAIADWVPDQIY 1006
Db 874 TDGQFLDSVIQNGYAFADRYDLGMSKANKYGTADQLVKAIKALHAKGLKVMADWVPDQWY 933
QY 1007 QLPKGEVTVATRTDAGAKRIADAIIDHSLYVANSKSSGDKYQAKYGEFLAEIKAKYDEM 1066
Db 934 TPFQKQEVVTVTRTDKFGKPIAGSIQNHSLYVTDTKSSGDDYQAKYGGAFDLDELKERYPEL 993
QY 1067 FKVNMISTCKPIDDSVKLKQWKAEEYFNGTNVLERGVYVLSDEATKYFTVTKEGNFIP 1126
Db 994 FTKKQISTQALDPSVKIKQWSAKYFNGSNILGRGADYVLSQVSNKYFNVASDITLFLPS 1053
QY 1127 QLTGKEKVITGFSDDGKIT-FTGSGTQAKSAFVTFNGNTYTFDARGHMVTNSEYSPNG 1185
Db 1054 SLGLK-VBESGIRYDGKGIYNSSATGTDQVKASFITEAGNLVYFGKGYMVTGAQ-TING 1111
QY 1186 KDVIYFLPGLMISLNAFYIDANGNTLYNSKQMGYKGTFTDVBSTDKGKESKVKVPR 1245
Db 1112 AN-YFLENGTALRNTIYTDAGNSHYANDGKRYENGYQQF-----GND-----WR 1157
QY 1246 YFTNEGVMAGVTVIDGFTQYFGEDEGFOAKDK-LVTFKGTYYFDDAHTGNGIKDWTM-RNI 1303
Db 1158 YP-KDGNMAGVLTVDGNTVQYFDKGVQVQAKDKLIVTRDGKRYFDQHNAAATNTFIADK 1216
QY 1304 NGKWWYFDANGVAATGAQVINGQKLYFNBEDSGQVGVVKNADGTYSKYKEGFGELVTNE 1363
Db 1217 TGHWYVYLGKDVAVTGAQTVGKLYFEANGQVQKGFVTSDEGKLYFYDVSQDMWTD 1276
QY 1364 PFTTGNVWYIYAGANKTWTGAQVINGQHLVFNADSGQVGVVKNADGTYSKYNASTGE 1423
Db 1277 FTEDKAGNMYFLGKDGAAVTGAQTIRGQKLYFKANGQVQKGVIVKGTGDKIRYYDAKSGE 1336
QY 1424 RLITNE-----PFTTG 1433
Db 1337 QVFNKTVKAADGKTYVINGDVAVDPVSVVKGTQFDKASGALRPNLKGQVLTGSGMYETA 1396
QY 1434 DNNWYIYGANGSVTGEVIGIGDDTVFFAKDGKVGQVQVTSAGNKRISYYYIGDSGKRAVST 1493
Db 1397 NHDWYI-QSGKALTGEQTINGQHLVYFKEDGHQVQKQVLTGTDGKRYVYDANGSQAFNK 1455
QY 1494 WIEIQGVTVYFDPKNGLA 1511
Db 1456 SVTVANGKTY-YFGNDGTA 1472
```

RESULT 14
ADX37280

ID ADX37280 standard; protein; 1590 AA.

```
XX ADX37280;
XX AC
XX DT
XX 21-APR-2005 (first entry)
XX DE Streptococcus sobrinus glucan binding protein B.
XX KW immunogenicity; immune stimulation; glucan binding protein-B;
XX KW microparticle; major histocompatibility complex; tooth disease.
XX OS Streptococcus sobrinus.
XX PN US2005031633-A1.
XX PD 10-FEB-2005.
XX PF 09-MAR-2004; 2004US-00797821.
XX PR 13-APR-1998; 98US-0081550P.
XX PR 08-JAN-1999; 99US-0115142P.
XX PR 12-APR-1999; 99US-00290049.
XX PR 07-MAR-2002; 2002US-0363209P.
XX PR 08-AUG-2002; 2002US-0402483P.
XX PR 07-MAR-2003; 2003US-00383930.
XX (SMIT/) SMITH D J.
XX PA (TAUB/) TAUBMAN M A.
XX PI Smith DJ, Taubman MA;
XX WPI; 2005-151644/16.
XX PT New composition comprising a fragment of a glucan binding protein-B
XX PT (GbpB) that binds to MHC Class II protein, and a biocompatible
XX PT microparticle, useful for producing an antibody (claimed) for immunizing
XX PT mammals against dental caries.
XX PS Claim 7; SEQ ID NO 37; 73pp; English.
XX CC The invention relates to a composition comprising a fragment of a glucan
XX CC binding protein-B (GbpB) and a biocompatible microparticle, where the
XX CC fragment binds to a major histocompatibility complex (MHC) class II
XX CC protein. The composition is useful for producing an antibody for
XX CC immunizing mammals against dental caries. This sequence corresponds to a
XX CC Streptococcus sobrinus GbpB protein of the invention.
XX SQ Sequence 1590 AA;
XX Query Match 41.9%; Score 3325.5; DB 9; Length 1590;
XX Best Local Similarity 45.0%; Pred. No. 1.2e-169;
XX Matches 710; Conservative 239; Mismatches 456; Indels 173; Gaps 33;
QY 1 MENKHYKLHKYKKQWVIAVASA-LATVLGGLSVTSSVSADETDQKTVTQSNSGTTA 59
Db 1 MEKNYRFXMHYKRWVTLVSAATMLASALG-----ASVASAD-----TDTASDD---- 46
QY 60 SLVTSPEATKEADKRTNTKEADVLTPAKETNAVEATTTNTQATAEAATTTATTADVAVAA 119
Db 47 -----SNQAVVTGDTTNNQATD-----QTSATAAT-----SQSASTDAATDQASAA 90
QY 120 VPKNEAVVTTDAPVTEKAEQPAVKAENVN-----TEVKAPEALKDSEVEEAL 171
Db 91 EQTQGTASTDTAAQTNTTANNE-AKWVPTENENQGTDEMLAEAKNVATAESDIPSDLA 149
QY 172 SLKNTKNDKGYKYVYNGDGHENKFAITVNGQLLYFGKDGALYTSSTYFTGTTNNVDG 231
Db 150 KMSNVKQVDGKYIYDQDGNVKNFAVSGDKIYFYFDETKYKTSKVDADRSSSAVSQN 209
QY 232 ---FSINNRAVDSSEASPELIDGYLTADSWPEPASTIKDGVTMQASTAEDFRPLLMWWP 288
Db 210 ATIFANNRAVSTSAKNFEAVDNYLTADSWYRPKSLKDGKTWBSGKDDFRPLLMWWP 269
QY 289 NVDTQVNYLVNMSKVFNLDKYSSTDQKQETLKVAAKDIQIKIEQIAEKSTQWLRETIS 348
```


Qy	1244	FRYFTNEGVMAGVTVIDGFTQYFEGEDGOAKDK-LVTTFKGTYYYDHAHTNGIKOTW-R	1301
Db	1158	WRYP-KDGNMVAELTVDGNVVQFYFDKGVOAKDKIIVTRDGKVRYFDQHNGNAVTTFFIA	1216
Qy	1302	NINGKWYYFDANGAATGAQVINGOKLYFNEDGSQVKGVVKNADGTSKYKEGFGELVT	1361
Db	1217	DKTGHWYYLKGOGVAVTGAQTGVGKQLXYPEANGQQVKGFVTSDEGLKPYDVDSDGMWT	1276
Qy	1362	NEFFTDDGNVWYAGANGKTVTGAAOVINGOHLVFNAEDGSQVKGVVKNADGTSKYKNAST	1421
Db	1277	DTIEBKAGNWFLGKDGAATVGAQTIRGKLXYFKANGQQVGDI VKGTGKI RYYDAKS	1336
Qy	1422	GERLTNE-----DFTS-----	1431
Db	1337	GEQVFNKTKAADGKTYVI GNDGVA VDPS VVKGTQTFK DAS GALR FYNLKQL VTGSGWE	1396
Qy	1432	TGDNNWNYLGANGSKVTGEVKLGDDTTYFAKODGQVKGGT V SAGNRI SVYYGDSGKRAV	1491
Db	1397	TANHDWYI-QSGKALTGEOTINGOHLVFKKHQVKGQQQLVTGTGDKRVYDANSDDQAF	1455
Qy	1492	STWIEIQPGVYVYDFDNGLA	1511
Db	1456	NKSVTVNGKTY-YFGNDGTA	1474

Search completed: February 11, 2006, 19:15:26
Job time : 159.097 secs

Qy	172	SLKNIKNIDGYYIYNVEDGGSHKENFALTUNGQLLYFGKDQALTSSTSYSTFTPGTTNIVDG	231
Db	150	KMSNVKQVDGYYIYDQDGNVKNFVSGVDKIYYFDGTGAYKDOTSKVDADKSSASQN	209
Qy	232	---PSINNRAYDSSEASFELIDGLYTADSWYRPASIIKOGVTWQASTAEDRPLLLMAWP	288
Db	210	ATTPAANNRAYSTSAKFEADVNTLTADSWYRPKSIKDGKTWTESGKDDFRPLLLMAWP	269
Qy	289	NVDTQVNLVYMSKVFNLDAKYSSTDKQETLKVAAKDIQIKIEQIOAEKSTOWLRETTIS	348
Db	270	DTETKRYVNVYMNKVGIDKTYTAEITSQADLTAAAEVLQARIEQKITSENNTKWLREAIS	329
Qy	349	AFVKTQPMWKENETEN--YSKGGEBHLOGGALLLYVND--RTPWMSDYRRLLNRATNQTGT	406
Db	330	AFVKTQPMNGESEKPY-----DDHLQNGALLFDNQDTPDTOSNYRLLNRTPTNQTGS	384
Qy	407	IDKSILDEQSDPNHMGGDFLLANDVLSNPVQAEQLNOIHLYLMWGSIVMGDKDANFD	466
Db	385	LDSRFTYNNDP--LGGYDFLLANDVNSNPVQAEQLNLHLYLLNFGSIYANDADANFD	442
Qy	467	GIRVDAVNDVADMLQLYTNFREYYGNKSEANALAHISVLEAWSLNDNHYNDKTDCAA	526
Db	443	SIRVDAEDNDVADQLQISSDYLKAAYGIDKNNKNNHNVSIVEAWSNDTPYLHDDGNL	502
Qy	527	LAMENKORLALLFSLAKPIKERTPAVSPLYNNTPTNTTQRDEKTDWINKDGSKAYNEDGTV	586
Db	503	MNMNKPRLSNLWLSAKPTDVR--GLNPLIHNSLVDEVDNR-----EVETV	548
Qy	587	KQSTIGYNEKYGDASGNYVPIRAHDNNVODIIAEIIKKEINPKSDGFTTIDAEKQAPE	646
Db	549	P-----SYGFARAHDSVQDIDRDIIKABINPNSFGYSTQBEIDQAFK	592
Qy	647	IYNKDMLSDKKYTLNNIIPAAYAAMLQNMETITRVYVYGLYDITDDGHYMETKSPYDTIVN	706
Db	593	IYNEDLKSDKKYTHYNVPLSYTLNLLTNKGSIPRVYVYGDMTDDGQYMAKNTVNYDAIES	652
Qy	707	LMKSRIKYVGGQARSYMLPTDGKMDNSVLYRTNVEVTSVRYGKDIINTANDTEGSKY	766
Db	653	LLKARMKYVAGGQAMQNYQI-----GNGEILTSVRYGKGLKQSD-KGDAT	697
Qy	767	SRTSGQVTLVANNPKMLDOSAKLNVEMGKIHANQKTRALIIVGTADGINKFTSDADIAA	826
Db	698	TRTSGYGVVNGNQNFSLDGKV--VALMMGAHANQEVYRALVMVSTKDGAVATYADADASKA	756
Qy	827	GYVKETDSNGLVFGANDIKGYETFDMSGFVAVVWPVPGASDNODIRVAPSTEAKKEGELT	886
Db	757	GLVKRTDENGLYFLNDDLKGAVNPQVSGFLQVWPVPGAADQDQIRVAASDSTASTDGK-S	815
Qy	887	LKATEAYDSQLIYEGFSNFOTIPDGSFPSVYTNRKIAENVDLPKSGWGTSEFEMAPQVSA	946
Db	816	LHQDAAMDVRMFESEFNSQSF--ATKEBEYTNVVIANNVDKFSWGIITDFEMAPQVSS	873
Qy	947	DDGTFLDSVTQNGYAFADRYDLAWSKNKYKSGKEDLBDALKALHKAGIOAIADWPDPQIY	1006
Db	874	TDCQFODSVTQNGYAFTRDYDLGMSKANKYGTADQLVKAIKALHAKGLKVMADWPDPQY	933
Qy	1007	QLPCKEVTYATRDGAGRKTIADAILIHSLSLVANSKSSGKDYQAKYGEFELAEKAKYPEM	1066
Db	934	TFPKQEVVTVTRDKFKPIAGSQIHNLSLYVTDPTKSSGDDYQAKYGGAFLELKEKYPEL	993
Qy	1067	FKYNMISTGKPIODDSVKLKQWKAEYFNGTNVBERGVYLSDBATGKYFVTVKEGNFIPL	1126
Db	994	FTKKQWSTGQALDPSVKIKQWSAKYFNGSNILGEGADYVLSQVSNKYFNVAADTFLPS	1053
Qy	1127	QLTGCKEVIITFGSDGKGITY-FGTSGTQAKSAFVTFNNGNTYYFDARGHMVTNSYSPNG	1185
Db	1054	SLLGK--VWESGIRYDGGKYVYNSSATGDQVKASFITEAGNLVYFGDGYVWVTAQ--TING	1111
Qy	1186	KDYTRFLPNCIMLSNAPYIDANGNTYLYNSKGQWY--KGGYTKFPDVSETDKDKESKVKV	1243
Db	1112	AN--YFLENGTALRNTIYTDQGNSHYANDGKRYENENYQGF-----GND-----	1157

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 19:39:30 ; Search time 33.2355 Seconds
(without alignments)
3776.130 Million cell updates/sec

Title: US-10-797-821-40
Perfect score: 7928
Sequence: 1 MENKHYKLHKYKQWVTIA.....PGVYVFDKNGLAYPPRVLN 1518

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/prodata/1/aaa/5 COMB.pep.*
2: /cgn2_6/prodata/1/aaa/6 COMB.pep.*
3: /cgn2_6/prodata/1/aaa/H COMB.pep.*
4: /cgn2_6/prodata/1/aaa/PCTUS COMB.pep.*
5: /cgn2_6/prodata/1/aaa/RE COMB.pep.*
6: /cgn2_6/prodata/1/aaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3502.5	44.2	1430	2 US-09-008-172-2	Sequence 2, Appli
2	3502.5	44.2	1430	2 US-09-210-361-6	Sequence 6, Appli
3	3502.5	44.2	1430	2 US-09-740-274-6	Sequence 6, Appli
4	3196.5	40.3	1475	2 US-09-007-999-2	Sequence 2, Appli
5	3196.5	40.3	1475	2 US-09-210-361-2	Sequence 2, Appli
6	3196.5	40.3	1475	2 US-09-740-274-2	Sequence 2, Appli
7	3155	39.8	1577	1 US-08-793-824-2	Sequence 2, Appli
8	3070	38.7	1375	2 US-09-210-361-4	Sequence 4, Appli
9	3070	38.7	1375	2 US-09-740-274-4	Sequence 4, Appli
10	2327	29.4	2057	2 US-09-499-203-2	Sequence 2, Appli
11	2093	26.4	1278	2 US-09-604-957-3	Sequence 3, Appli
12	2093	26.4	1781	2 US-09-995-749A-2	Sequence 2, Appli
13	1584.5	20.0	545	2 US-09-604-957-4	Sequence 4, Appli
14	1560.5	19.7	545	2 US-09-995-749A-10	Sequence 10, Appli
15	1262.5	15.9	523	2 US-09-604-957-5	Sequence 5, Appli
16	1262	15.9	522	2 US-09-995-749A-11	Sequence 11, Appli
17	1220.5	15.4	535	2 US-09-604-957-7	Sequence 7, Appli
18	1220.5	15.4	535	2 US-09-995-749A-13	Sequence 13, Appli
19	1090	13.7	584	2 US-09-604-957-6	Sequence 6, Appli
20	1088	13.7	584	2 US-09-995-749A-12	Sequence 12, Appli
21	586.5	7.4	349	2 US-09-009-620-2	Sequence 2, Appli
22	523	6.6	2710	1 US-08-480-604A-6	Sequence 6, Appli
23	523	6.6	2710	1 US-08-405-496A-6	Sequence 6, Appli
24	523	6.6	2710	1 US-08-915-136-6	Sequence 6, Appli
25	523	6.6	2710	2 US-08-957-310-6	Sequence 6, Appli
26	523	6.6	2710	2 US-10-011-366-6	Sequence 6, Appli
27	523	6.6	2710	2 US-09-084-517-6	Sequence 6, Appli

28	447	5.6	811	1 US-08-480-604A-7	Sequence 7, Appli
29	447	5.6	811	1 US-08-405-496A-7	Sequence 7, Appli
30	447	5.6	811	2 US-08-915-136-7	Sequence 7, Appli
31	447	5.6	811	2 US-08-957-310-7	Sequence 7, Appli
32	447	5.6	811	2 US-10-011-366-7	Sequence 7, Appli
33	447	5.6	811	2 US-09-084-517-7	Sequence 7, Appli
34	447	5.6	812	1 US-08-480-604A-29	Sequence 29, Appli
35	447	5.6	812	2 US-08-915-136-29	Sequence 29, Appli
36	447	5.6	812	2 US-09-084-517-29	Sequence 29, Appli
37	436.5	5.5	866	2 US-09-545-773-2	Sequence 2, Appli
38	436.5	5.5	866	2 US-10-222-038-2	Sequence 2, Appli
39	379.5	4.8	1231	2 US-08-714-741-41	Sequence 41, Appli
40	366	4.6	2366	1 US-08-480-604A-10	Sequence 10, Appli
41	366	4.6	2366	1 US-08-405-496A-10	Sequence 10, Appli
42	366	4.6	2366	2 US-08-915-136-10	Sequence 10, Appli
43	366	4.6	2366	2 US-08-957-310-10	Sequence 10, Appli
44	366	4.6	2366	2 US-10-011-366-10	Sequence 10, Appli
45	366	4.6	2366	2 US-09-084-517-10	Sequence 10, Appli

ALIGNMENTS

RESULT 1
US-09-008-172-2
; Sequence 2, Application US/09008172
; Patent No. 6127602
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE REFERENCE: 0358D
; CURRENT APPLICATION NUMBER: US/09/008,172
; CURRENT FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-008-172-2

Query Match	44.2%;	Score	3502.5;	DB 2;	Length	1430;	
Best Local Similarity	48.1%;	Pred. No.	4.7e-222;	Mismatches	434;	Gaps	32;
Matches	724;	Conservative	247;	Mismatches	434;		
y	1	MENKHYKLHKYKQWVTIAVASVALATVLGGLSVTTSSVSAD---	ETQDKTWTQ-SNSG	56			
b	1	METKERYKMHKVKHWHVTVAVAS-GLITL--GTTTLGSSVSAETEQQTS	DKVVTQKSEDD	57			
y	57	TTASLVTSPKATKADKRTNTEADVLTPAKETNAVEATTNTTQATABAAT-	ATTADV	115			
b	58	KAASESSQTDAPKTKQAQTEQTAQ-----SQANVADTSTSIKETPSQNI	TQANSDDK	112			
y	116	AAVAPNKEAVVTTDAPAVTTTEKAEQPAVKAEVNVTEVKA-----PEAL	KDSEVEA	169			
b	113	TVNTKSEEAQTSB-----RTKQSEEAQTASSQAL--TQAKAELTKQR	TAAQENKNPVD	167			
y	170	ALSLANKINIDGKYYVNVNEDSGHKNFAITVNGQLLYFGKD--GALT	SSSTYSFFPTGTTNI	228			
b	168	LAAPINVKQIDGKYYIIGSDGPKKFNALTVNNKVLKFKNTGALT	DTTSQYFKQGLTKL	227			
y	229	VDGFSINNRAVDSSSEAPELIDGYLTADSWRYPASIIKDGVTWQAST	AEFRPLLMWMP	288			
b	228	NNDYTPHNQIVNFENTSLETIDNYVTADSWRYPKDKILKNGKWT	WTASSESLRPLMSWMP	287			
y	289	NVDVTQVNLNYSKV-FNLDAKYSTDKQETLKVAAKDIIQIKI	EOKIOAEKSTOWLRETI	347			
b	288	DKQGTQIAYLNNYNNQOGLGTGENYTTADSSQESLNLAQT	VQVKIETKISQTOQTOWLRDII	347			
y	348	SAFVKTQPMNKETENYSKGGEDHLQGGALLVYVNDSTRTPWANS	SDYRRLNRTATNTQGTI	407			


```
QY 742 TN--EYVTSVRVYKDIWNTANDTEGSKYSRTSGOVTLVANNPKLNDQSAKLNVEMCKIHA 799
Db : : : : :
QY 783 LINGEIMSSVRVYKGAEEANQL-GTAEITRNOGMLVLTANRPDMKGLANDRLVNNGAHK 841
Db : : : : :
QY 800 NOKYRALIVGTADGIGNFTSDADAIAAGVYKETSNGVLTFCGANDIKGYETFDMSGFVAV 859
Db : : : : :
QY 842 NOAYRELLLSKSTGLATYIKDS-VPAGLVRYTDNQGNTFTTADDIAGHSTVEVSGYLAV 900
Db : : : : :
QY 860 WYPVGASNDQDIRVAPSTBAKEGELTLKATEAYDSQLIYEGFSNFQTIIPDGSDPSVYTN 919
Db : : : : :
QY 901 WYPVGASENQDARTKAS--STKKEQVPESSAALDSQVIYEGFSNFQDFV--KTPSQYTN 956
Db : : : : :
QY 920 RKIAENVDLFKSGVTSFEMAFQFUSADGDTGLDSVIQNGYAFADRYDIAMSKNNKYGSK 979
Db : : : : :
QY 957 RVIAQNAKLFKEGITSFEPFQVYSSQDGTLDLSIENGYAFEDRYDIAMSKNNKYGL 1016
Db : : : : :
QY 980 EDLRDALKALHKAGIOAIADWPDQIYQLPGKEVVTATRTDAGRKIADAIIDHSLYVAN 1039
Db : : : : :
QY 1017 KDLMDALRALHABGISAIADWPDQIYNLPGKEVVTASRTNSYGTTPRPAEINSLYAAK 1076
Db : : : : :
QY 1040 SKSSGKDYQAKYGGEFELAKKAKYPMFKVMNMISTGKPIDDSVYKLQWKAEBYFNGTNVLE 1099
Db : : : : :
QY 1077 TRTEGNDFOGKYGGAFDELDELKAKYPAI FERVOLISNGRKLTTNEKITOWSAKYFNGSNIQ 1136
Db : : : : :
QY 1100 RGVGYVLSDEATGKYFTVYKGNFIPQLTGKEKVTGFSGDKGHITFGTSGTOAKSAP 1159
Db : : : : :
QY 1137 TGAARYVLQDNATNQYFSVRAGOTFLPKQMT--EITGSGFRRVGDDVQYLSIGLYLAKNTF 1194
Db : : : : :
QY 1160 VTFNGNT--YFEDARGHMVTNSEYSPNGKDVYRPLNGIMLSNIFYIDANGNTLYNSKG- 1217
Db : : : : :
QY 1195 IQVGANQWYTFDKNGNMVTEGEVIDGK--YFPLDNGQLRHLVROSGDHVYYPDKGV 1252
Db : : : : :
QY 1218 QMYKGYTKFDVSETDKGSKSVKVFYFTFTEGVNAKGVTVIDGFTQYFGE--DGFQAKD 1276
Db : : : : :
QY 1253 QAFNGFY-----DFAGPRQDV---RYFDGNGQMYRGLHDMYGTTFYFDEKTIQAKD 1301
Db : : : : :
QY 1277 KLVT-P-KGYTYFPDAUTGN-GTKDWTNRNNGK-WYFDPDANGVAATGAQVINGOKLYFNED 1333
Db : : : : :
QY 1302 KTRFADGRTRYFIPDTGNLAVNRFAQNPENKAWYILDSNGYAVTGLQTINGKQYFDNE 1361
Db : : : : :
QY 1334 GSOVKGVVKNADGTSYKVEGELVNTNEFFTTDGNVWYAGANGKTVTGAQVINGOHL 1393
Db : : : : :
QY 1362 GRQVGHFTVNNQRY--FLDGSGEIAPSRFVTENKKNYVYDGNKLVKGAQVINGNHY 1419
Db : : : : :
QY 1394 YFNADGSQVKGGVVKNADGTSYKYNASTGERLTNEFFTTGDNWYIYGANGKSVTGEVKI 1453
Db : : : : :
QY 1420 YFNNDYSQVKGAW--ANGRY--YDGDGSGQAVSNQFIQIAANQWAYLNDGHKVTGLQNI 1474
Db : : : : :
QY 1454 GDDTYFPAKDGKQVKGQTVSAGNGRISYVYDGSGRKRAVSTWIEIQPGVVVYFDKGLA 1511
Db : : : : :
QY 1475 NKKVYFGSNGAQVKGKLLTV-QGKKCYFDAHTGEQVNRVFEAARGCWYFNSAQOA 1531
Db : : : : :
```

RESULT 8

```
US-09-210-361-4
; Sequence 4, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
```

```
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-210-361-4

Query Match      38.7%; Score 3070; DB 2; Length 1375;
Best Local Similarity 46.2%; Pred. No. 1.5e-193;
Matches 661; Conservative 223; Mismatches 412; Indels 134; Gaps 33;

QY 1 MENKHYKLHKVKKQWVTIAVASVALATVGLSLVTSVSSADETQDK--TVTQNSGTT 58
Db : : : : :
QY 59 ASLVTSPKATKADKRTWKTEADVLTPAKETNAVETATTTTQATAEAAATATTATDAVA 118
Db : : : : :
QY 53 ASLVTTSAAKETLTATDTSTATSATSOPTATVTDNVSTTN-QST---NTTANTANFVVK 108
Db : : : : :
QY 119 AVPNKEAVVTTDA-PAVTKEAEQ-----PATVKAEEVNVTEVKAP-- 159
Db : : : : :
QY 109 PTTTSEQAKTDMSDKIITTSKAVNRLTATGKFPANNNTAHPKTVTDKIVPIKPKIGK 168
Db : : : : :
QY 160 --AALKDSEVEAALSLKNIKNIDGKYVYVNEDEGSHKENFAITVNGQLLYFGDKGALTSS 217
Db : : : : :
QY 169 QPSSLSQDDIALGNVKNIRKVNKYVYKEDGTLOKYNALNINKTFFDFETGALSNTT 228
Db : : : : :
QY 218 TVSFPTGTNN--IVDGFSSINNRAYDSSEASPELIDGYLTADSWYRPASIIKDGVTWQAST 275
Db : : : : :
QY 229 LPSKKGNTNNDNTNSFAQYNQVYSTDVANFEHVDHYLTAEWSYRPKYILKDKGTWQST 288
Db : : : : :
QY 276 AEDFPEPLMAWPNVDVTQVNYLVNYSKVPNLDAKYSTSDTKQETLKVAANDIQIKIEQIK 335
Db : : : : :
QY 289 EKDFPEPLMTWPDQETQRQYVYVNAQIGHOTVNTATSPQLNLAAQTIOTKIEBKIT 348
Db : : : : :
QY 336 AEKSTQWLRETSIAFVKTPQPMNKETENYSKGGEDHLQGGALLVYVNDNR--TPWANSYDR 394
Db : : : : :
QY 349 AEKNTNLQRTLSIAFVKTPQSAWNSDSEK---PFDDHLQKGLLYSNNSKLSQANSYR 404
Db : : : : :
QY 395 RLNRATATQGTGIDKSIILDEQSDPNH-----MGGEFFLLANDVLSNPVVAEQLNQIHY 449
Db : : : : :
QY 405 ILNRTPTNQTG-----KKDPRTADRTTIGGYEFLILANDVNSNPVVAEQLNWLHP 455
Db : : : : :
QY 450 LANNVGSIVMGDKDANFDGIRVDVAVDNDADMLQLYTNYFREYVYVGNKSEANALAHISVLE 509
Db : : : : :
QY 456 LNNFNIYANDPDANFDSIRVDVAVDNDADLLQIAGDYLKAAKGIHNDKAANDHUSILE 515
Db : : : : :
QY 510 AWSLNDNHNVDKTDGAALAMENKQRLALLFLSLAKPIKERTPAVSPLYNNTF--NTTQDEBK 568
Db : : : : :
QY 516 AWSYNDTPYLHDDGDGNMINDNRLSLLSLYSLAKPLNQRS--GWNPLITNSLVNRTDDNAE 574
Db : : : : :
QY 569 TDWINKGSKAYNEDGTQVKQSTIGKNEYKGDASGNVYFIRAHNNVQDIIAIIKKEIN 628
Db : : : : :
QY 575 T-----AAVPSYSFIRAHDSVQDLIRNIIRTEIN 604
Db : : : : :
QY 629 PKSDGFTITDAEMKQAFELYNKMDLSSDKKYTLNNIIPAAAYVLMQMETITRVYVYDGLYT 688
Db : : : : :
QY 605 FNVVGSYFTTEIKKAFELYNKMDLATEKKYTHYNTALSYALLTNKSSVPVRYGDMFT 664
Db : : : : :
QY 689 DDGHYMETKSPYYDTIVMLMKSRIKYVSGGQQRSYWLPDTDGKMDNSDVELYRTNEVYTS 748
Db : : : : :
QY 665 DDQYMAHKTINYEALETLLKARIKYVSGGQAWNQ-----QVGNSS-----EIIIS 710
Db : : : : :
QY 749 VRYGKDIIMTANTEGSKYSRTSGOVTLVANNPKLNDQSAKLNVEMCKIHAQKYRALIV 808
Db : : : : :
QY 711 VRYGKALKATDT-GDRTTTRTSQVAVIEGNNPSLRKASDRVVVNNVNGAAHQNQAYRPLLL 769
Db : : : : :
QY 809 GTADGINKPTSDADALAAAGYVKETDSNGVLTGANDIKGYETFDMSGFVAVVWVPVGCASDN 868
Db : : : : :
```


Db 770 TTDNGIKAYHSDQE--AAGLVRYTNDRGELIFTAADIKGYANPQVSGYLGVWVPVGAAD 827
Qy 869 QDIRVAPSTBAKKEGELTLKATEAYDSQIYEGFSNFQITPDGSDPSPVYNNKIAENVLD 928
Db 828 QDVRVAASTAPSTDGK-SVHQNAALDSRVWFEFSNFQAF--ATKKEEYTNVVIKNDVK 884
Qy 929 FKSQWVTSFEMAPQFVSADGTFLDSVIQNGYAPADRYDLAMSKNNKIYKSKEDLRDALKA 988
Db 885 FAEWGVTFDEMAPQVYSSSTGSLDSVIQNGYAFTRDYDLGISKPNKYGTADDLVKAICA 944
Qy 989 LHKAGIOAIADWPDQIYQIPGKEVVTATRTDAGRKIADAIIDHSLYIVANSKSSGKDIQ 1048
Db 945 LHSKGIKVMADWPDQIYALPEKEVVTATRTDAGRKIADAIIDHSLYIVANSKSSGKDIQ 1004
Qy 1049 AKYGGFBLAELKAKYEMFKNMISTGKPIDDSVKLKWKAIEYFNGTGNVLERGVGVYLS 1108
Db 1005 AKYGGFLELOAKYELFARKQISGVPMDFPVSKIKWSAKYFNGTGNVLERGVGVYLS 1064
Qy 1109 EATGKYFTVTKGNFPLQLT---GKEKVTGFSDDGKGYITFGTSGTOAKSAFVTFNG 1164
Db 1065 QATNTYFSLVSDNTFLPKSLVNPNGHTSSVTGLVPDGGKYVYVYSTSGNQAQNAFISLGN 1124
Qy 1165 NTYYEDARGHMTNSEYSPNGKDVYRFLPNGIMLSNAFYIDANGTYL--YNSKGQMYKG 1222
Db 1125 NWYIFDNGYMTGAQ-SINGANY-FLSNGIQLRNAY--DNGNKVLSYNGDGRYEN 1180
Qy 1223 GYTKFQVSETDKGSKSVKVPYFTNEGVMAGVTVTIDGFTQYFGEDEGQAKDLV-TF 1281
Db 1181 GYLYF-----GQQ-----WYFON-GIMAVGLTRVHGAQVYFASGFQAKGQFITTA 1226
Qy 1282 KGKTYFPAHTNGIKIDTW-RNNGKMYFEDANGVAATGAQVNGKQYFNEDEGSOVKG 1340
Db 1227 DGKLYFDRDSGNQISNRFVNSKGEWFLPDHNGVAVTGVTFNGQRLYFKPKNVQAKGE 1286
Qy 1341 VKNADGTSKYKEGFGELVTNEFFTDGDNVWYAGANGKTVTGAOVING 1390
Db 1287 FIRDANGYLYRYPDPSNGEYRNRFRVNSKGEWFLPDHNGIATVGTARVWNG 1336

RESULT 9
US-09-740-274-4
; Sequence 4, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-4

Query Match 38.7%; Score 3070; DB 2; Length 1375;

Best Local Similarity 46.2%; Pred. No. 1.5e-193;
Matches 661; Conservative 223; Mismatches 412; Indels 134; Gaps 33;
Qy 1 MENKIHYKLVKKKQWVTIYAVASVALATVGLGSLVTTSSYASDETQDK--TVTQSNSTGT 58
Db 1 MEKKVFPKURKKKRWVTIYAS-AVVTL---TSLGSLVKADSTDRQQAQTESQ---- 52
Qy 59 ASLVTSPEATKEADKRTNTKEADVLTAPAKETNAVETATTTNTQATAEAATATTADVAVA 118
Db 53 ASLVTTSEAAKETLTATDTSTATSATSQPTATVTDVNSITTN-QST---NTTANTANFVK 108
Qy 119 AVNKEAVVTDA-PAVTTEKABEQ-----PATVKAENVNTEVKAPE-- 159
Db 109 PTTTSEQAKTDSDKIITTSKAVNRLTATGKFPANNTAHPKTVDTKVIPKPKIGKJK 168
Qy 160 --AALKSDSEVEALSUKNIKIDGKYVYVNEDEGSHKENFAITVNGQLLYFGKQALTS 217
Db 169 QPSSLSQDDIAALGNVKNIRKVGKYYKYKEDGTLQKNVALNINGKTFPFDDETGAUSN 228
Qy 218 TYSFTPTGTTN--IVDGFSINNRAYDSSEASFELIDGTLTADSWYRPAASIKDGVTWOAST 275
Db 229 LPSKKGNIINNDNTNSFAQYVQVYSTDVANFEHVDHLYTABSRYRKYILKDGKWTQST 288
Qy 276 AEDFRPLMAMPNDVDTQVNYLNMYSKVNFLDAKYSSTDQKQETLKVAAKDQIKIEBQIQ 335
Db 289 EKDFRPLMTWPDQETQRYVYNNYAQLGIHOTYNTATSPQLQLNAAQTQIKIEKIT 348
Qy 336 AEKSTOMARETISAFVKTQPOWNETENYSGGEGDHLOGGALLYNDYR-TPWANSDYR 394
Db 349 AEKNTNMLRQTISAFVKTQSANNSDSEK----PFDDHLQKGLLYNNNSKLTQANSNYR 404
Qy 395 RLNRATNTGTTIDKSIILDEQSDPNH---MGGPDFLLANDVDLSNPVVQAQLNIOHY 449
Db 405 ILNRFTNQTG-----KKDPRYTABRTIGGYEFLLANDVDNSNPVVQAQLNWLHP 455
Qy 450 LNMWGSIVMGDKDANPDGIRVDADVNDADMLQLYTNYPREYGVNKSANALAHISVLE 509
Db 456 LNMFGNIYANDPANFDSIRVDADVNDADLLQIAGDYLKAAKGIHKNDKCAANDHLSILE 515
Qy 510 ANSLADNHNKDTGDAALAMENKQRLALLFSLAKPIKERTPAVSPLYNNTF-NTTORDEK 568
Db 516 AWSYNDTPYLHDDGDNMINMDNRLRLSLYSLAKPLNQRS-GMNPLITNSLVNRTDQNAE 574
Qy 569 TDWINKGSKAYNEDGTQVKTGTYGKNEYGASGNYVFIRAHNNVDQIIAIEIKKEIN 628
Db 575 T-----RAVPSYFIRAHDSVQDLIRNIITEIN 604
Qy 629 PKSDGFTITDAEMKQAFIYNKMDLSSDKKYTLNNIPAAAYAVMLQNMETITRYVYDLYT 688
Db 605 PNWGVSYFTTEBIKKAFIYNKDLLATEKKYTHYNTALSYALLTNKSSVPRVYVYDGMFT 664
Qy 689 DDGHYMETKSPYDITVNLMSRIKYVSGGQARSYWLPTDGMKMDNSDELRYTNEVYTS 748
Db 665 DDGQYMAHKTINYEAITELLKARIKYVSGGQARNQ-----QVGNs-----BIITS 710
Qy 749 VRYGKDIIMTANDTEGSKYTSRTSQVTLVANNNKLNLDQSAKLNVEMGKTHANOKYRALIV 808
Db 711 VRYGKALKATDT-GDRTTTSQVAVIEGNNSLRLKASDRVVVNNGAHKQAYRPLLL 769
Qy 809 GTADGINKFTSDADAIAAGYVKETDSNGVLTFCANDIKGYETDFDMSGFVAVVVPVGAASN 868
Db 770 TTDNGIKAYHSDQE--AAGLVRYTNDRGELIFTAADIKGYANPQVSGYLGVWVPVGAAD 827
Qy 869 QDIRVAPSTBAKKEGELTLKATEAYDSQIYEGFSNFQITPDGSDPSPVYNNKIAENVLD 928
Db 828 QDVRVAASTAPSTDGK-SVHQNAALDSRVWFEFSNFQAF--ATKKEEYTNVVIKNDVK 884
Qy 929 FKSQWVTSFEMAPQFVSADGTFLDSVIQNGYAPADRYDLAMSKNNKIYKSKEDLRDALKA 988
Db 885 FAEWGVTFDEMAPQVYSSSTGSLDSVIQNGYAFTRDYDLGISKPNKYGTADDLVKAICA 944
Qy 989 LHKAGIOAIADWVPDQIYQIPGKEVVTATRTDAGRKIADAIIDHSLYIVANSKSSGKDIQ 1048

Db 945 LHSKGIKWADVPDMYALPEKEVVTATRVDKYCTPPVAGSQIKNTLYVVDGKSSGKDOQ 1004
QY 1049 KYGGSEFLAELKAKYEMFKVNMISTGKPIDDSVKLKQWKABYFNGTNNVLERGVGVLSGD 1108
Db 1005 KYGGAFLEELQAKYPELFPARKQISTGVPMDPSPVKIKQWSAKYFNGTNTLGRGAGYVLKD 1064
QY 1109 EATGKYFTVTKEGNFTPLOLT---GKEKVIITGSSDGKGIYFGTSGTQAKSAFVTFNG 1164
Db 1065 QATNTYFSLVSNTFLPKSLVNPNGHTSSVGLVDPGKGYYVYTSNGQAKNAFTLGN 1124
QY 1165 NTYFDFARGHMTNSEYSPNGKDVYRFLPENGMLSNFAYIDANGTYL---YNSKGMQYKG 1222
Db 1125 NWYFDDNGYMTGAQ-SINGANY-FLSNGIQLRNLAY--DNGKVLSSYNGDGRYEN 1180
QY 1223 GYTKFVDSSTDGKESKVVKRYFTNEGMAKGVTVIDGFTQYFGEDEGQAKDLV-TF 1281
Db 1181 GYVLF-----GQQ-----WRYFON-GIMAVGLTRVHGAVQYFDASGFAQAGQFITTA 1226
QY 1282 KGKTYFDDAHTNGIKDTH-RNINGKMYFFDANGVAATGAOVINGOKLYFNEDEGSOVKGG 1340
Db 1227 DGKLYFDDRDSGNQISNRFRVNSKGEWFLFDHNGVAVTGTVTFNGQORLYFKPENGVOAKGE 1286
QY 1341 VKNADGTSKYKKGEGELVTNEFTTDCGNVMYAGANGKTVTGAQVING 1390
Db 1287 FIRDANGILRYDPSNGEVRNRFVRNSKGEWFLPDHNGIATGARVYNG 1336

RESULT 10
US-09-499-203-2
; Sequence 2, Application US/09499203
; Patent No. 6570065
; GENERAL INFORMATION:
; APPLICANT: KOSSMANN, Jens
; APPLICANT: WELSH, Thomas
; APPLICANT: QUANZ, Martin
; APPLICANT: KNUTH, Karola
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternansucrase
; FILE REFERENCE: 147-196P
; CURRENT APPLICATION NUMBER: US/09/499,203
; CURRENT FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2057
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-09-499-203-2

Query Match 29.4%; Score 2327; DB 2; Length 2057;
Best Local Similarity 38.2%; Pred. No. 2.9e-144;
Matches 577; Conservative 209; Mismatches 509; Indels 217; Gaps 49;

QY 144 ATVKAERVNTEVKAPPAALKDSEVEAALS-----LKNKNIDGKYYVN 187
Db 300 AVANTEIINGKLQGRDA--NGNQVNAFSDKVAGNTFFFDANGVMLTGLQISGKTYILD 357
QY 188 EDGSHKENFAITVNGQLLYFGKD-GALTSSSTYSFPPGTNTINVDGFSINNRAYDSEASF 246
Db 358 EQGHLRKNYAGTFNNGFMFYFADTGAGKTAIEYQPDQGLVSSQSNENTPHNAKSYDKSSF 417
QY 247 ELIDGVLTDMSVYRPASIIKDGVTWQASTAEDFRPLLMAMWPNVDQVNYLYNM-SKVPN 305
Db 418 ENVDGYLTADTWYRPDILKNGDWTWASTETDMRPLMTWPDQOTQANYLNFMSKGLG 477
QY 306 LQAKYSSTDKQETLKVAAKDIOIKIEQIOAKESTOWLRETTISAFVKTOPQWNKETENYS 365
Db 478 ITTYTAASQKTLNDAAVFIQTALIQOISLKSTEWLEDAIDSFVKTOANWKNQTEDEA 537
QY 366 KGGEDHLOGGALLYVND-S-RTPWANS-DYRRLNRTATNTQGTIDKSIILDEQSDPNHMG 423
Db 538 FDGLQ-WLQGGFLAYQDDSHRTPNTDSGNNRKLGRQPIN-----IDGSKDTTDDGK 587
QY 424 FDFELLANDVLSNPVQAQLNQIHYLMNWSGSIWGDKNDFDGIIRVDAVDNDADMLQI 483

Db 588 SEFLLANDIDNSNPVQAQLNQLNHLNMFSSITGNNDNANPDGIRVDAVDNDADMLKI 647
QY 484 YTNYPREYGVNKSANALAHISVLEAWSLNDNHNNDKTDGAALAMENKQRLALLFSLAK 543
Db 648 AGDYFKALYGTDKSDANANKHLSILEDWNGKDPQVYVNOQNAQLTWDY-----695
QY 544 PIKERTPAVSPLNYNTFTTQORDEKTDW-----INKDSKAYNEDGTQVKQSTIGKN 595
Db 696 -----TVTQFGNSLTHGANNRNNMYFLDTGYLYLNGDLNKKITVDKNNRPNSGTLVNRI 748
QY 596 EKYGDAS--GNVYFTRAHDNNVQDII--AEIIKKEINPKSDQFTITDAEMKQAFELYNKD 651
Db 749 ANSGDTKVIPNYSFVRAHDYDAQDPIRKAMIDHGIIKNQDFTTFD--QLAOCMEFYKD 806
QY 652 M--LSSDKKTYLNNIPAAAYAVMLQMETITRYYGDLTDDGHYMETIKSPYYDTIVNLMM 709
Db 807 QENPSGPKYNDYLPAYAMLLTNKDTVPVYVGYDMYLEGGQYMEKGTIYNPVSALLK 866
QY 710 SRIKTVSGGQQRSTWLPDGDHMONSDVELYRTNEVYTSVRYGKDIIMANDT---EGSKY 766
Db 867 ARIKTVSGGTQMAT---DSSGK-DLKDGE---TDLTTSVREGKIMTSDQTTQDQNSOD 918
QY 767 SRTSQVTLVANNPKLNDQSAKLNVEGKIHANOKYRALIVGTADGIKNFTSDADIAA 826
Db 919 YKQIGIVGVGNPDKLNDKTIITLHMGKAHKNQYRALVLSNDSGDIDVDYDDKAPT 978
QY 827 GYVKETSDNGVLTFGAN-----DIKGYETFTDMSGFVAVVVPVVGASDNQDIR- 872
Db 979 ---RTNDNGDLIFHKTTNFVKDGTIINYEKMGSLNALISGLGVWVPGVSGSDQART 1034
QY 873 VAPSTEAKKEBELTLKATEAVDSQLIYEGFSNFQIPDGSDPVSVTNKRKIAENVDLFKSW 932
Db 1035 VATESSSSDNGSV-FHNSAALDSNVIYEGFSNFQAMP--TSPEQSTNVVVIATKANLFKSL 1091
QY 933 GYVTSPEMAPOQFVSADDG---TFLDSVIQNGYAFADRYDLAMSKNN-----KYGSKED 981
Db 1092 GITSFELAPQVRSSGDTNYGMSFLDSFLNNGYATFDYDLGFKADGNPNPKYGTQDQ 1151
QY 982 LRDALKALHKGAGIAIADWPDIYQYLPKQEVVATRTDAGRKTADAIIDHSLYVANSK 1041
Db 1152 LRNAIEALHKGQMAIADWPDIYALPKQEVVATRVDERGNQLKDTDFVNLLYVANTK 1211
QY 1042 SSGKDYQAKYGGEFLEAELKAKYEMFKVNMISTGKPIDDSVKLKQWKABYFNGTNNVLERG 1101
Db 1212 SSGVDYQAKYGGEFLEAELKAKYEMFKVNMISTGKPIDDSVKLKQWKABYFNGTNNVLERG 1271
QY 1102 VGVLSDATGKYFTVTKEGN-FIPLQLTGKEKVIITGFSDDGKGIYFGTSGTQAKSAFV 1160
Db 1272 AYYVLKDWATNQYFNIAKTNEVFLPLQONKD-AQTGFISDASGVKYYSISGYQAKDTFI 1330
QY 1161 -TFNGNTYFDFARGHMTN-----EYSPNGKDVYRFLPENGMLSNFAYIDANG 1209
Db 1331 EDGNGMYVYFDKGYMVRSSQOGENPIRTVETSVNTRNGNYFPMNGVELRKGFGTDNSGN 1390
QY 1210 TYLYNSKQMYKGGYTKFDVSETDKDKESKVVKRYFTNEGMAKGVTVIDGFT-QYFG 1268
Db 1391 VYFPDQOGKVRDKYINDAN-----NFIHLNVDDGTMSRGLPKFSDTLQYFA 1438
QY 1269 EDGFOAKDLV-TFKGKTYFDDAHTGN--GIKDTWRNINGKMYFF---DANGVAATGA 1320
Db 1439 SNGVQIKDSYAKDSKNGKYFDSATGNNDTGKAQTH---DNGYIYITDSDANNITGVNT 1495
QY 1321 QVINGQKLYFNBEG---SQVKGGVVKNADG-----TYSKYKEGFGELVTFNEFTTQD 1369
Db 1496 DYTAVITSSLEDGLPANAPYGVVTKDQNGNDLKWQYINHHTKQYEGQQVQVTRQYTDKSG 1555
QY 1370 NWV---YYAG-----ANGKTVTGAQVIN-----GOHLYFNADGSOVKG- 1404
Db 1556 VSWNLITFAGDLOGQRLWDSRALTPFTKTMNQISFISYANRNDGLFLNAP-YGVKGY 1614
QY 1405 -----GV--VKNADGTVSKYNAS---TGERLTNEFTTGNNNYYIG- 1441

Db 1615 QLAGMSNOYKGOQVTTIAGVANVSCKWLSLISFNGTQVWIDSQALNTNFTHDMNQKVFVNT 1674
Qy 1442 -ANGKSVTGEVKIGDDTYFFPAKDGKQVKQGVTSAGNGRISYYGDSGKRAVSWIE-IQP 1499
Db 1675 TSULDGLFLNAPYRQPGYKLAGLAKNYYNTVTS-----QQYFDDQG-----TWSQVVLG 1726
Qy 1500 GVVYVFDKGLA 1511
Db 1727 GQTVVVDNHALA 1738

RESULT 11

US-09-604-957-3
; Sequence 3, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; PRIOR FILING DATE: 2000-06-28
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1278
; TYPE: PRT
; ORGANISM: Lactobacillus reuteri
US-09-604-957-3

Query Match 26.4%; Score 2093; DB 2; Length 1278;
Best Local Similarity 37.2%; Pred. No. 3.8e-129;
Matches 511; Conservative 197; Mismatches 414; Indels 250; Gaps 42;
Qy 18 TIAVASVALATVLGGLSVT---TSSVSADETQKTVTQNSGTTASLVTSPKATK----- 69
Db 54 TIKLTDQYQALNGLOVLLRSKAADGNPSGNTVTDQFSKRYATGTGDFVYKVGNGQ 113
Qy 70 -----EADKRTNKAD---VLTPAKETNAVETATTTNTQATAEATATTATTADV--AVA 118
Db 114 VEPFGWHATQSNKDSQMIIVLNGKE---VKRLQVNDTK---EGAAGFNRNDVYKVP 167
Qy 119 AVPNK-----EAVVITDAVATTEKABEQPATVKAHV-----NTEVKAPER----- 160
Db 168 AIENSSMSFGQGIITL-----PVTYKNNVQLVHRFSNDVKTGEGNYVDFWS 214
Qy 161 ---ALKDSVEAALSILK--NIKNIDGKYVYVN-EDGSHKENFAITVNGQLLYFGKD-GAL 213
Db 215 ELPVVDKSKQNGPLKQGLQTINGQQYIDPTGQPRKFLQLSQNNWIYFDSGTGVG 274
Qy 214 TSSSTYSFPTGTNIVDGFSSINNAYDSSEAFELIDGYLTADSWTRPASIIKDGVTWQA 273
Db 275 TWALELOFAGKGTVSSNEQYRNGNAAVSYDDKSIENVNGVYLTDWYRPKILKDGTTWTD 334
Qy 274 STAEPRPLMAWPNVDVQVNYLNTMSK-----VFNLDAKYSSTDQOETLKVA 323
Db 335 SKETDMRPLTMYWPNPTLTQAYLYNTMKQGNLLPSALPFFNADA-----DPAE-LNHYS 388
Qy 324 KDQIKIEOKIQAESTQWLRETISAFVKTPQWNKETENYSKGGEDHLOGGALLYND 383
Db 389 EIVQONIEKRISETGNDWRLMDHDFVTNNPMNKDSNVNPSGIQ---FQGGFLKYENS 446
Qy 384 SRTPWANSYRRLNRTATNQTGTIDKSIILDEQSDPNHMGGFDFLLANDVLSNPVVAEQ 443
Db 447 DLTPTYANSYRLLGRMPIN-----IKDQT-----YEQQFLLANDIDNSNPVVAEQ 493
Qy 444 LNIHIYLMNWSGIVGDKDANFDGIRVADVNDVMDLQLYTNYFREYYGVNKSANAL 503
Db 494 LNWLYLLNFGTITANNQANFDSVRVDPADNIDADMNIADQYFNAAVYGMDS-DAVSNNK 552

Qy 504 HISVLEAWSLNDHYNDKTDGAALAMENKORLALLFSLAKPIKERTPAVSPLYNNTFTNT 563
Db 553 HINILEDWHADEYFNKIGNPOLTWDDTIKNSLHGLS----- 591
Qy 564 QRDEKTDWINKOGSKAYNEDEGTVKQSTIGKYNEKYGD-ASGNVYFIRAHNNVQDIIAEI 622
Db 592 -----DATNRWGLDAI VHQSLADRENNSTENVVIPNYSFVRAHDNNSQDQIQNA 640
Qy 623 IKKEINPKSDGFTITDAEMKQAFEIYNKDWLSDDKYYTLNNIPAAVAMLONMETITRVY 682
Db 641 I-RDVTGK-DYHTFTFDEBQGDIDAYIQDNSTVKKYNLYNIPASVAILTNDKDTIPRVY 698
Qy 683 YGDLTYDDGHYMETKSPYYDTIWNLMKSRIKYVSGQAORSYWLPTDGMKDNDSVLEYLT 742
Db 699 YGDLTYDDGGQYMEHQRYVDTLTNLLKSRVYVAGGQSMQT-----MSVGGN 745
Qy 743 NEVYTSVRYGKDIIMTANDTEGSKYSRTSQVTLVANNPKLNLDQSAKLANVEMGIHANQK 802
Db 746 NNILTSVRYGKGAMTATDT-GTDETRTQIGVGVVSNTPNLKLGVDNKVVLHMGAAHKQKQ 804
Qy 803 YRALIVGTADGIGNFTSDADATAAGVYKETSNGVLTFGAND-----IKGYET 850
Db 805 YRAAVLTITDGVINTYSDQAPVA-----MTDENGDLYLSSHNLVNVNGKEADTAVQGYAN 860
Qy 851 FDMGSGFVAVVVPVVGASDNODIRVAPSTEAKKEGELTKATEAYDSOLIVEGFSNFQITPD 910
Db 861 PUVSGYLAWVVPVVGASDNODARTAPSTE-KNSGNSAYRTNAAPFDSNVIFAEFNFVYTP- 918
Qy 911 GSDPSVYTNRKIAENVDLFKSMGVTSFEMAPQVFSADDTGLDSDVIONGVYAFADRYDLAM 970
Db 919 -TKESERANVRIAQNADPFASLGFTSFEMAPQVNSKSDRTFLDSTIDNGYAFDTRDYLGM 977
Qy 971 SKNKYKSGKEDLRDALKALHAKAGIAIADWDPDQIYOLPGKEVTVATRTDAGRKIADAI 1030
Db 978 SEPKNYGTDEDLRNAIQALHAKAGLQVMAWDPDQIYNLPFGKEVATVTRVDRGNVWKDAI 1037
Qy 1031 IDHSLYVANSKSGKDYQAKYGGEBFLAELKAKYPEMFKNVNMISTGKPIDDSVKLKQWKA 1090
Db 1038 INNLYVNTVITGG-EYQKYGAFDLQKLYPEIFTKKQVSTGVAIDPSQKITEWSAK 1096
Qy 1091 YFNGTNVLERGVYVLSDEATGKYF---TVTKEGNFIPQLTGKKEKVIITGFSDDGKITY 1147
Db 1097 YFNGTNILHRGSGYVLKADG-QQYVNLGTTTQK--FLPQLTGK-- 1139
Qy 1148 FGTSGTQAKSAFTVN-GNTYTFDARGHVMVTSNPSGKDVYRFLPNGIMLSNAPYIDA 1206
Db 1140 -----QGNFVKGNDGNYFYDLAGNV-----KNTFIEDS 1171
Qy 1207 NGNTYLYNSKGQMYKGYTKPDVSETDKDGKSKVVKPRYFTNEGVMKGVTVIDGF--- 1263
Db 1172 VGNWTF-----DQDQ---KMWENKHFVD-----VDSYGEK 1199
Qy 1264 -TQYFEGEDGFOAKDLVTFKGTYYFADATNGNGIKDTRNNGKWKYFYFDANG 1314
Db 1200 GTYFELKNGVSRGGLVQTDNGTYFYFDNY-GHGVENQITINAGAMYTLTDENG 1250

RESULT 12

US-09-995-749A-2
; Sequence 2, Application US/09995749A
; Patent No. 6867026
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: BO43388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28

QY 600 DASGNVFIIRAHNNVQDIIEIIKKEINPKSDGFTITDAEMKQAFIYNKQMLSSDKKY 659
Db 154 --MANYIFIRAHDSVQTVIAKIIKAQINPKTDGLTFLDELKQAFIYNEDMRQAKKY 211
QY 660 TLANNIPAAVAVMLQNNMETITRVYGYDLYTDDGHYMETKSPYDTIVNLMKSRIKYSGGQ 719
Db 212 TQSNIPATAVAMLSNKSITRLYGYDMSYSDGQYMATKSPYYDAITLLKARIKYAAGGQ 271
QY 720 AQRSYMLPTDGMNDSDVELYRTNEVYTSVRYGKDIINTANDTEGSKYRSRTSGQVTLVANN 779
Db 272 DMKITTV--EGDKSHMD--WDVTGVLTSVRYGTGANEATD-QGSEATKQGMVITSNN 325
QY 780 PKLNLQDSAKLVNEMGKIHANOKYRALIVGTADGINKFTSDADAIAGYVKTDSNGVLT 839
Db 326 PSLKLNQNDKVIIVNMGAAHKNQYRPLLLTTKGLTSYTS--DAAAKSLYRKTNDRGELV 898
QY 840 FGANDIKGYETFMGSGFVAVVVPVPGASDNQDIRVAPSTEAKKEGELTLKATEAYDSOLII 899
Db 384 FDASDIQGLYNLPQVSG--LAVVVPVPGASDNQDIRVVAASNKANATGV--YESSSALDSOLII 442
QY 900 EGFNSFQT-IPDGSFSPVYTNRKIAENVDLFKSWGVTSPFEMAPQFVSADGTFLDSVION 958
Db 443 EGFNSFQDFVTNKDSD--YTNKKIAQNVQLFKSWGVTSPFEMAPQYVSSSDGSLDSIIQ 499
QY 959 GYAFADRYDLAMSKNNKYGSKEDLRDALKALHKGAGIQAIDWVPDQ 1004
Db 500 GYAFEDRYDLAMSKNNKYGSKQODMINAVKALHKSGLQVIADWVPDQ 545

RESULT 14

US-09-995-749A-10

; Sequence 10, Application US/09995749A

; Patent No. 6867026

; GENERAL INFORMATION:

; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA

; APPLICANT: DIJKHUIZEN, LUBBERT

; APPLICANT: RAHAOUI, HAKIM

; APPLICANT: LEER, ROBERT-JAN

; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES

; FILE REFERENCE: BO43388-CIP

; CURRENT APPLICATION NUMBER: US/09/995,749A

; CURRENT FILING DATE: 2001-11-29

; PRIOR APPLICATION NUMBER: 09/604,957

; PRIOR FILING DATE: 2000-06-28

; PRIOR APPLICATION NUMBER: EPO 00201871.1

; PRIOR FILING DATE: 2000-05-25

; NUMBER OF SEQ ID NOS: 19

; SOFTWARE: Patent in Ver. 2.1

; SEQ ID NO 10

; LENGTH: 545

; TYPE: PRT

; ORGANISM: Streptococcus mutans

US-09-995-749A-10

Query Match

Best Local Similarity 19.7%; Score 1560.5; DB 2; Length 545;

Matches 316; Conservative 99; Mismatches 121; Indels 51; Gaps 12;

QY 427 LLANDVDLSNPVQAEQLNQIHYLMNWSIVMGDKDANFDGIRVDAVDNVDMLQLYTN 486
Db 1 LLANDIDNSNPVQAEQLNWLHYLMNYSIVANDPEANFDGVRVDAVDNVDNADLLQIASD 60
QY 487 YFREYGVNKSANALAHISVLEAWSLNDNHNNDKTDGAALAMENKORLALLFSLAKPIK 546
Db 61 YLKAHGVDSKSEKAINHLSLEAWSNDNPQYNKDTKGALPDKLRLSLLYALTRPUE 120
QY 547 -----ERTPAVSPLYNNFTNTQDEKTDWINKDGSKAYNEDGTGVKQSTIGKYNKYG 599
Db 121 KDAASGNKIRSGLEPVITSLN-----NRSABGNKSER-- 153
QY 600 DASGNVFIIRAHNNVQDIIEIIKKEINPKSDGFTITDAEMKQAFIYNKQMLSSDKKY 659
Db 154 --MANYIFIRAHDSVQTVIAKIIKAQINPKTDGLTFLDELKQAFIYNEDMRQAKKY 211

QY 660 TLANNIPAAVAVMLQNNMETITRVYGYDLYTDDGHYMETKSPYDTIVNLMKSRIKYSGGQ 719
Db 212 TQSNIPATAVAMLSNKSITRLYGYDMSYSDGQYMATKSPYYDAITLLKARIKYAAGGQ 271
QY 720 AQRSYMLPTDGMNDSDVELYRTNEVYTSVRYGKDIINTANDTEGSKYRSRTSGQVTLVANN 779
Db 272 DMKITTV--EGDKSHMD--WDVTGVLTSVRYGTGANEATD-QGSEATKQGMVITSNN 325
QY 780 PKLNLQDSAKLVNEMGKIHANOKYRALIVGTADGINKFTSDADAIAGYVKTDSNGVLT 839
Db 326 PSLKLNQNDKVIIVNMGAAHKNQYRPLLLTTKGLTSYTS--DAAAKSLYRKTNDRGELV 898
QY 840 FGANDIKG-YETFDMSGFGVAVVVPVPGASDNQDIRVAPSTEAKKEGELTLKATEAYDSOLII 898
Db 384 FDASDIQGLYNLPQVSG--LAVVVPVPGASDNQDIRVVAASNKANATGV--YESSSALDSOLII 441
QY 899 YEGFSNFQT-IPDGSFSPVYTNRKIAENVDLFKSWGVTSPFEMAPQFVSADGTFLDSVIO 957
Db 442 YEGFSNFQDFVTNKDSD--YTNKKIAQNVQLFKSWGVTSPFEMAPQYVSSSDGSLDSIIQ 498
QY 958 NGYAFADRYDLAMSKNNKYGSKEDLRDALKALHKGAGIQAIDWVPDQ 1004
Db 499 NGYAFEDRYDLAMSKNNKYGSKQODMINAVKALHKSGLQVIADWVPDQ 545

RESULT 15

US-09-604-957-5

; Sequence 5, Application US/09604957

; Patent No. 6486314

; GENERAL INFORMATION:

; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA

; APPLICANT: DIJKHUIZEN, LUBBERT

; APPLICANT: RAHAOUI, HAKIM

; APPLICANT: LEER, ROBERT-JAN

; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN

; FILE REFERENCE: BO 43388

; CURRENT APPLICATION NUMBER: US/09/604,957

; CURRENT FILING DATE: 2000-06-28

; PRIOR APPLICATION NUMBER: 00201871.1

; PRIOR FILING DATE: 2000-05-25

; NUMBER OF SEQ ID NOS: 17

; SOFTWARE: Patent in Ver. 2.1

; SEQ ID NO 5

; LENGTH: 523

; TYPE: PRT

; ORGANISM: Leuconostoc mesenteroides

US-09-604-957-5

Query Match

Best Local Similarity 15.9%; Score 1262.5; DB 2; Length 523;

Matches 269; Conservative 84; Mismatches 170; Indels 55; Gaps 8;

QY 427 LLANDVDLSNPVQAEQLNQIHYLMNWSIVMGDKDANFDGIRVDAVDNVDMLQLYTN 486
Db 1 LLANDIDNSNPVQAEQLNWLHYLMNYSIVANDPEANFDGIRVDAVDNVDNADLLQIASD 60
QY 487 YFREYGVNKSANALAHISVLEAWSLNDNHNNDKTDGAALAMENKORLALLFSLAKPIK 546
Db 61 YKFLAYGVQDNDATANHLSLEAWSNDNPLVYTDGSGNQLTMDYVHTQLIWSLTSSD 120
QY 547 ERTPAVSPLYNNFTNTQDEKTDWINKDGSKAYNEDGTGVKQSTIGKYNKYGDSGNV 606
Db 121 IR-----GTMQRFVDYVMDRSDNSTENE-----ALPNYS 150
QY 607 FRAHDNNVDIIABIIKKEINPKSDGFTITDAEMKQAFIYNKQMLSSDKKYTLNIPA 666
Db 151 FVRAHDSVQTVIAQIVSDLYPDVENS LAPTEQLAAAFKYNEDKADKKYQYNMAS 210
QY 667 AYAVMLQNNMETITRVYGYDLYTDDGHYMETKSPYDTIVNLMKSRIKYSGGQARSYWL 726
Db 211 AYAVMLLTNKDTVPVRYGYDLYTDDGQYMATKSPYYDAITLLKARQYVYAGGQSM----- 265

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 20:39:23 ; Search time 135.87 Seconds
(without alignments)
4668.162 Million cell updates/sec

Title: US-10-797-821-40
Perfect score: 7928
Sequence: 1 MENKHYKLHKVKQWVTIA.....PGVYVFDKNGLAYPPRVLN 1518

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA_Main:*
1: /cgn2_6/prodata1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/prodata1/pubpaa/US08_PUBCOMB.pep:*
3: /cgn2_6/prodata1/pubpaa/US09_PUBCOMB.pep:*
4: /cgn2_6/prodata1/pubpaa/US10A_PUBCOMB.pep:*
5: /cgn2_6/prodata1/pubpaa/US10B_PUBCOMB.pep:*
6: /cgn2_6/prodata1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7928	100.0	1518	4	US-10-383-930-40
2	7928	100.0	1518	5	US-10-797-821-40
3	3818	48.2	1554	4	US-10-383-930-38
4	3818	48.2	1554	5	US-10-797-821-38
5	3502.5	44.2	1430	3	US-09-740-274-6
6	3502.5	44.2	1430	4	US-10-383-930-36
7	3502.5	44.2	1430	5	US-10-797-821-36
8	3325.5	41.9	1590	4	US-10-383-930-37
9	3325.5	41.9	1590	5	US-10-797-821-37
10	3196.5	40.3	1475	3	US-09-740-274-2
11	3196.5	40.3	1475	4	US-10-383-930-34
12	3196.5	40.3	1475	5	US-10-797-821-34
13	3070	38.7	1375	3	US-09-740-274-4
14	3070	38.7	1375	4	US-10-383-930-35
15	3070	38.7	1375	5	US-10-797-821-35
16	2804	35.4	1365	4	US-10-383-930-39
17	2804	35.4	1365	5	US-10-797-821-39
18	2795.5	35.3	1497	5	US-10-484-218-18
19	2557.5	32.3	1595	5	US-10-484-218-20
20	2327	29.4	2057	4	US-10-417-280A-2
21	2142.5	27.0	1777	5	US-10-484-218-12
22	2093	26.4	1781	3	US-09-995-749A-2
23	2084	26.3	1006	5	US-10-484-218-22
24	2024.5	25.5	1771	5	US-10-484-218-14
25	1560.5	15.7	545	3	US-09-995-749A-10
26	1262	15.9	522	3	US-09-995-749A-11
27	1220.5	15.4	535	3	US-09-995-749A-13

28	1140.5	14.4	787	5	US-10-484-218-16	Sequence 16, Appl
29	1088	13.7	584	3	US-09-995-749A-12	Sequence 12, Appl
30	1055.5	13.3	525	5	US-10-484-218-23	Sequence 23, Appl
31	545.5	6.9	224	5	US-10-484-218-4	Sequence 4, Appl
32	541	6.8	223	5	US-10-484-218-10	Sequence 10, Appl
33	523	6.6	2710	4	US-10-011-366-6	Sequence 6, Appl
34	523	6.6	2710	4	US-10-354-774-6	Sequence 6, Appl
35	523	6.6	2710	4	US-10-271-012-6	Sequence 6, Appl
36	523	6.6	2710	4	US-10-729-123-6	Sequence 6, Appl
37	523	6.6	2710	4	US-10-729-039-6	Sequence 6, Appl
38	523	6.6	2710	5	US-10-729-527-6	Sequence 6, Appl
39	523	6.6	2710	5	US-10-727-898-6	Sequence 6, Appl
40	523	6.6	2710	5	US-10-728-696-6	Sequence 6, Appl
41	523	6.6	2710	6	US-11-001-241-6	Sequence 6, Appl
42	501	6.3	223	5	US-10-484-218-6	Sequence 6, Appl
43	447	5.6	811	4	US-10-011-366-7	Sequence 7, Appl
44	447	5.6	811	4	US-10-354-774-7	Sequence 7, Appl
45	447	5.6	811	4	US-10-271-012-7	Sequence 7, Appl

ALIGNMENTS

RESULT 1
US-10-383-930-40
; Sequence 40, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 40
; LENGTH: 1518
; TYPE: PRT
; ORGANISM: Streptococcus salivarius
US-10-383-930-40

Query Match	100.0%;	Score 7928;	DB 4;	Length 1518;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1518;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MENKHYKLHKVKQWVTIAVASVALATVGLGLSVTTSSVSADETQDKTIVTQNSGTTAS	60	
Db	1	MENKHYKLHKVKQWVTIAVASVALATVGLGLSVTTSSVSADETQDKTIVTQNSGTTAS	60	
QY	61	LVTSPTEAKDEADKRTNTKEADVLTPAKETNAVETATTTTQATAEAATTTATTADVAVA	120	
Db	61	LVTSPTEAKDEADKRTNTKEADVLTPAKETNAVETATTTTQATAEAATTTATTADVAVA	120	
QY	121	PNKEAVTTDPAPVTTKEAEPATVKAEEVNVNTEVKAPEAAKDEVEAALSLKIKNID	180	
Db	121	PNKEAVTTDPAPVTTKEAEPATVKAEEVNVNTEVKAPEAAKDEVEAALSLKIKNID	180	
QY	181	GKYYVNEGSHKENFALTNGQLLYFGKDGALTSSSTYSFTPGTNTIVDGSINNAYD	240	
Db	181	GKYYVNEGSHKENFALTNGQLLYFGKDGALTSSSTYSFTPGTNTIVDGSINNAYD	240	
QY	241	SSEASFELIDGLTADSWYRPASIIKDGVTWQASTAEDFRPILMAWPNVDQVNYLNYM	300	
Db	241	SSEASFELIDGLTADSWYRPASIIKDGVTWQASTAEDFRPILMAWPNVDQVNYLNYM	300	
QY	301	SKVFNLDKYSTDKQETLKVAAKDQIKIEQIKQAEKSTQWLRETIISAFVKTQPNWKE	360	
Db	301	SKVFNLDKYSTDKQETLKVAAKDQIKIEQIKQAEKSTQWLRETIISAFVKTQPNWKE	360	

QY 361 TENYKGGGDEHLQGGALLYVNDSTRTPWANSYRRLNRTATNTQGTIDKSIILDESDPNH 420
DB 361 TENYKGGGDEHLQGGALLYVNDSTRTPWANSYRRLNRTATNTQGTIDKSIILDESDPNH 420
QY 421 MGGFDFLLANDVDLSNPVQAOLQIHYLMNWGSI VMGDGDANPDGIRVDADVNDADM 480
DB 421 MGGFDFLLANDVDLSNPVQAOLQIHYLMNWGSI VMGDGDANPDGIRVDADVNDADM 480
QY 481 LQLYTNYFREYYGVNKSEANALAHISVLEAWSLNDNHNDKTDGAALAMENKORLALLFS 540
DB 481 LQLYTNYFREYYGVNKSEANALAHISVLEAWSLNDNHNDKTDGAALAMENKORLALLFS 540
QY 541 LAKPIKERTPAVSPLYNNTFNTQDEKTDWINKDGSKAYNEDGTVKQSTIGKYNEKYD 600
DB 541 LAKPIKERTPAVSPLYNNTFNTQDEKTDWINKDGSKAYNEDGTVKQSTIGKYNEKYD 600
QY 601 ASGNVYFIRAHNNVDIIAEIIEKINPKSGDFTITDAEMKQAFIYNKMLSSDKKYT 660
DB 601 ASGNVYFIRAHNNVDIIAEIIEKINPKSGDFTITDAEMKQAFIYNKMLSSDKKYT 660
QY 661 LNNIPAAVAMLQNMETITRVYVYGDLYTDDGHYMETKSPYYDTIVNLMKSRIKYYSGGQA 720
DB 661 LNNIPAAVAMLQNMETITRVYVYGDLYTDDGHYMETKSPYYDTIVNLMKSRIKYYSGGQA 720
QY 721 QRSYMLPTDCKMNSDVELYRNEVYTSVRYGKDINTANDTEGSKYSRTSGQVTLVANNP 780
DB 721 QRSYMLPTDCKMNSDVELYRNEVYTSVRYGKDINTANDTEGSKYSRTSGQVTLVANNP 780
QY 781 KLNLDQSAKLVEMGKIHANQYRALIVGTADGINKFTSDADAIAGYVYKETSNGVLTF 840
DB 781 KLNLDQSAKLVEMGKIHANQYRALIVGTADGINKFTSDADAIAGYVYKETSNGVLTF 840
QY 841 GANDIKGYETFDMSGFVAVVVPVPGASDNQDIRVAPSTEAKGEGELTKATEAYDSQLIYE 900
DB 841 GANDIKGYETFDMSGFVAVVVPVPGASDNQDIRVAPSTEAKGEGELTKATEAYDSQLIYE 900
QY 901 GFSNFQTIIPDGDSPSYVTRKIAENVDLFPKSGVGTSEFMAPOFVSADDDGTFDLSVITQNGY 960
DB 901 GFSNFQTIIPDGDSPSYVTRKIAENVDLFPKSGVGTSEFMAPOFVSADDDGTFDLSVITQNGY 960
QY 961 AFADRYDLAMSKNKGKXEDLRDALKALHKAGIOAIADWPDQIYQLPKGKVVATRTD 1020
DB 961 AFADRYDLAMSKNKGKXEDLRDALKALHKAGIOAIADWPDQIYQLPKGKVVATRTD 1020
QY 1021 GAGRKIADAIIDHSLVYANSKSGKDYQAKYGEEFLAELKAKYPEMFVKNMISTGKPIDD 1080
DB 1021 GAGRKIADAIIDHSLVYANSKSGKDYQAKYGEEFLAELKAKYPEMFVKNMISTGKPIDD 1080
QY 1081 SVKLLQWKAEYFNGTNVLERGVGYVLSDEATGKYFTVTKEGNFIPLOLTGKEKVIITGFS 1140
DB 1081 SVKLLQWKAEYFNGTNVLERGVGYVLSDEATGKYFTVTKEGNFIPLOLTGKEKVIITGFS 1140
QY 1141 DGKGIITYFGTSGTQAASAFVTNGNTYFDDARGHVMVNTSEYSPNGKDVRFLPNGIMLSN 1200
DB 1141 DGKGIITYFGTSGTQAASAFVTNGNTYFDDARGHVMVNTSEYSPNGKDVRFLPNGIMLSN 1200
QY 1201 AFYIDANGNTYLYNSKGQMYKGGYTKFDVSETDKDGKESKVVKFRYFTNEGVMAGKVTVI 1260
DB 1201 AFYIDANGNTYLYNSKGQMYKGGYTKFDVSETDKDGKESKVVKFRYFTNEGVMAGKVTVI 1260
QY 1261 DGFTQYFGSDGFOAKDLVTFPGKTYTFDAHTGNGIKOTWRNINGKWIYFDANGVAATGA 1320
DB 1261 DGFTQYFGSDGFOAKDLVTFPGKTYTFDAHTGNGIKOTWRNINGKWIYFDANGVAATGA 1320
QY 1321 QVINGOKLYFNEDEGSQVKGKVNADGTSYKKEGFGELVTNEFFTTDGNVWVYAGANGK 1380
DB 1321 QVINGOKLYFNEDEGSQVKGKVNADGTSYKKEGFGELVTNEFFTTDGNVWVYAGANGK 1380
QY 1381 TVTGAQVINGQHLVFNADGSQVKGKVNADGTSYKYNASTGERLTNEFFTTGDNWNYYI 1440
DB 1381 TVTGAQVINGQHLVFNADGSQVKGKVNADGTSYKYNASTGERLTNEFFTTGDNWNYYI 1440

QY 1441 GANGKSVTGEVKIGDDTYFFFAKDGKQVKGQTVSAGNGRISYYYGDSGKRAVSTWIEIQPG 1500
DB 1441 GANGKSVTGEVKIGDDTYFFFAKDGKQVKGQTVSAGNGRISYYYGDSGKRAVSTWIEIQPG 1500
QY 1501 VVYVFDKNGLAYPPRVLN 1518
DB 1501 VVYVFDKNGLAYPPRVLN 1518
RESULT 2
US-10-797-821-40
; Sequence 40, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 40
; LENGTH: 1518
; TYPE: PRT
; ORGANISM: Streptococcus salivarius
US-10-797-821-40
Query Match 100.0%; Score 7928; DB 5; Length 1518;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1518; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MENKHYKLHKVKKQWVTIAVASVALATVGLGSLVTTSSVSADETQDKTVTQSNSTTAS 60
DB 1 MENKHYKLHKVKKQWVTIAVASVALATVGLGSLVTTSSVSADETQDKTVTQSNSTTAS 60
QY 61 LVTSPEATKADKRTNTEADVLTPAKETNAVETATTTNTQATAEAATTTATTADVAVAV 120
DB 61 LVTSPEATKADKRTNTEADVLTPAKETNAVETATTTNTQATAEAATTTATTADVAVAV 120
QY 121 PNKEAVVTTDAPAVTTEKAEQPAVTKAEVNVNTEVKAPAEALKDSEVAALS.LKNIKID 180
DB 121 PNKEAVVTTDAPAVTTEKAEQPAVTKAEVNVNTEVKAPAEALKDSEVAALS.LKNIKID 180
QY 181 GKYYVNVNEDGSHKENPAITVNGQLLYFGKDGLTSSSTYSFTPGTNNIVDGSINNRAYD 240
DB 181 GKYYVNVNEDGSHKENPAITVNGQLLYFGKDGLTSSSTYSFTPGTNNIVDGSINNRAYD 240
QY 241 SSEASFELIDGVLTDADSVYRPASIIKDGVTWQASTAEDFRPLLMAWPNVDQVNYLNYM 300
DB 241 SSEASFELIDGVLTDADSVYRPASIIKDGVTWQASTAEDFRPLLMAWPNVDQVNYLNYM 300
QY 301 SKVFNLDAKYSSTDQETLKVAAKDIQIKIEQIKQAEKSTQWLRETI.SAFVKTQPQWNKE 360
DB 301 SKVFNLDAKYSSTDQETLKVAAKDIQIKIEQIKQAEKSTQWLRETI.SAFVKTQPQWNKE 360
QY 361 TENYKGGGDEHLQGGALLYVNDSTRTPWANSYRRLNRTATNTQGTIDKSIILDESDPNH 420
DB 361 TENYKGGGDEHLQGGALLYVNDSTRTPWANSYRRLNRTATNTQGTIDKSIILDESDPNH 420
QY 421 MGGFDFLLANDVDLSNPVQAOLQIHYLMNWGSI VMGDGDANPDGIRVDADVNDADM 480

Db 421 MGGDFELLANDVLSNPVQAEQLNQHILYLMNWSIVMGDKDANFDGIRVDAVDNDADM 480
QY 481 LQLYTYNFRYYGVNKSSEANALAHISVLEAWSLNDNHNDKTDGAALAMENKORLALLFS 540
Db 481 LQLYTYNFRYYGVNKSSEANALAHISVLEAWSLNDNHNDKTDGAALAMENKORLALLFS 540
QY 541 LAKPIKERTPAVSPLYNNTFTTQDEKTDWINKDGSKAYNEDGTQKSTIGKYNKYGD 600
Db 541 LAKPIKERTPAVSPLYNNTFTTQDEKTDWINKDGSKAYNEDGTQKSTIGKYNKYGD 600
QY 601 ASGNYVFIKRAHDNNVDIIABIIKKEINPKSDGFTTDAEMKQAFIYNKMDLSSDKYTT 660
Db 601 ASGNYVFIKRAHDNNVDIIABIIKKEINPKSDGFTTDAEMKQAFIYNKMDLSSDKYTT 660
QY 661 LNNIPAAVAMLQNMMETIRVYVYGDLYTDDGHYMETKSPYDIIIVNLMKSRIKYGSGGQA 720
Db 661 LNNIPAAVAMLQNMMETIRVYVYGDLYTDDGHYMETKSPYDIIIVNLMKSRIKYGSGGQA 720
QY 721 QRSYWLPTDGMNDSDVELYRTNEVYTSVRYGKDIINTANDTEGSKYSRTSGQVTLVANRP 780
Db 721 QRSYWLPTDGMNDSDVELYRTNEVYTSVRYGKDIINTANDTEGSKYSRTSGQVTLVANRP 780
QY 781 KNLNDOSAKLNVEMGKIHANOKYRALIIVGTADGINKFTSDADAIAAGYVKETDSNGVLTF 840
Db 781 KNLNDOSAKLNVEMGKIHANOKYRALIIVGTADGINKFTSDADAIAAGYVKETDSNGVLTF 840
QY 841 GANDIKGYETFDMSGFVAVVWVPVGASDNQDIRVAPSTEAKKEGELTLKATEAYDSQLIYE 900
Db 841 GANDIKGYETFDMSGFVAVVWVPVGASDNQDIRVAPSTEAKKEGELTLKATEAYDSQLIYE 900
QY 901 GFSNFOTIPDGDSPSVYNNRKAENVDLPKSGWGVTSFEMAPQVSDGTFDLSVIONGY 960
Db 901 GFSNFOTIPDGDSPSVYNNRKAENVDLPKSGWGVTSFEMAPQVSDGTFDLSVIONGY 960
QY 961 AFADRYDLAWSKNNKYGSKEDLRDALKALHKAGIQAIADWVPDQIYQLPGKEVVTATRD 1020
Db 961 AFADRYDLAWSKNNKYGSKEDLRDALKALHKAGIQAIADWVPDQIYQLPGKEVVTATRD 1020
QY 1021 GAGRIADIADIDHSLYVANSKSGKDYQAKYGEFELAEKAKYPENFKVNMISTGKPIDD 1080
Db 1021 GAGRIADIADIDHSLYVANSKSGKDYQAKYGEFELAEKAKYPENFKVNMISTGKPIDD 1080
QY 1081 SVKLAKQWKAEBYFNGTNVLRGVGYVLSDEATGKYFTVTKEGNFIPLQLTGKEKVIITGFS 1140
Db 1081 SVKLAKQWKAEBYFNGTNVLRGVGYVLSDEATGKYFTVTKEGNFIPLQLTGKEKVIITGFS 1140
QY 1141 DGKGYTFGTSQAKSAFVTNGNTYYFDARGHVMVTNSEYSPNGKDVYRFLPNGIMLSN 1200
Db 1141 DGKGYTFGTSQAKSAFVTNGNTYYFDARGHVMVTNSEYSPNGKDVYRFLPNGIMLSN 1200
QY 1201 AFYIDANGNTLYNSKGMQYKGYTKFDYSETDKDGESKVVKFRYFTNEGVMAGKVTYI 1260
Db 1201 AFYIDANGNTLYNSKGMQYKGYTKFDYSETDKDGESKVVKFRYFTNEGVMAGKVTYI 1260
QY 1261 DGFTQYFGEDGFQAKDKLVTFKGYTYFFDAHTGNGIKDTRWNINGKYYFDFANGVAATGA 1320
Db 1261 DGFTQYFGEDGFQAKDKLVTFKGYTYFFDAHTGNGIKDTRWNINGKYYFDFANGVAATGA 1320
QY 1321 QVINGOKLYFNEDGSGQVGVKNADGTYSKYKEGFGELVTFNEFTTGDGNVWYAGANGK 1380
Db 1321 QVINGOKLYFNEDGSGQVGVKNADGTYSKYKEGFGELVTFNEFTTGDGNVWYAGANGK 1380
QY 1381 TVTGAQVINGOHLXFNADSGQVKGKGVKNADGTYSKYNASTGBRLTNEFTTGDNNWYI 1440
Db 1381 TVTGAQVINGOHLXFNADSGQVKGKGVKNADGTYSKYNASTGBRLTNEFTTGDNNWYI 1440
QY 1441 GANGKSVTGEVKIGDDTYFFAKDQKQVKGQTSVAGNGRISYYYGSGKAVSTWIEIQPG 1500
Db 1441 GANGKSVTGEVKIGDDTYFFAKDQKQVKGQTSVAGNGRISYYYGSGKAVSTWIEIQPG 1500
QY 1501 VVYVFDKNGGLAYPPRVLN 1518
|||||

Db 1501 VVYVFDKNGGLAYPPRVLN 1518
RESULT 3
US-10-383-930-38
; Sequence 38, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 38
; LENGTH: 1554
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-383-930-38
Query Match 48.2%; Score 3818; DB 4; Length 1554;
Best Local Similarity 50.1%; Pred. No. 1.2e-198;
Matches 793; Conservative 215; Mismatches 434; Indels 130; Gaps 24;
QY 1 MENKIHYKLHKYKQWVTIIVASVALATVVLGGLSVTSSVSADETDQDKTVTQSNSTTAS 60
Db 1 MEKHLHYKLHKYKQWVTIIVASIGLVSLVG----AGTVSA---EDKVANDTTAQATVG 52
QY 61 LVTSPEATKEADKRTTKTEADVLTAKETNAVETAT-TTNTQATATAAATATTADVA-VA 118
Db 53 VDTGQDQATTNDANTNTTDTADQSANTNQDQAGSDQSNQDQAKQDATTANDRNOADNS 112
QY 119 AVPNKEAVVTTDAPAVTTEKAEQPAVKAEVVTNEVKAPEALAKDSEVEAALSLKNIKN 178
Db 113 QTDNQAQTDQATSPATDGTSGVORRDA---ANVATAADQEGQTAPSEQEKSAALSLDNVKL 169
QY 179 IDGKYVYVYVNEGDGSHKENFAITVNGQLLYFGKD-GALTSSSTYSFTPTGTTNIVDGFGINNR 237
Db 170 IDGKYVYVYVQADGSGYKKNFAITVNGQWLYFSDTGALSSSTYSFSQGTTLNLDVDFSSHNK 229
QY 238 AYDSEASAFELIDGVLTDADSWYRPASIIKDGVTWQASTABDPRPLLMANWPNVDVTQVYL 297
Db 230 AYDSTAKSPFELVNGYLTANSWYRPAGILRNGQTWEASNENDLRPVLMSWPPDKDTQVAYV 289
QY 298 NYMSKVFNL-DAKYSGSTDKQETLKVAADQIQIKIQIOAEKSTOWLRETISAFVKTOPQ 356
Db 290 NTMNYKLSANEVETNETSVDLNEKASIQTKIQKITSNDNSTOWLRTAMEAFVAAQPK 349
QY 357 MNKETENYSKGGEDHLOGGALLYVNDSTRTPWANSDYRLNRTATNQTGTIDKSIILDEQS 416
Db 350 WNMSTENFKG--DHLQGGALLYTNSDLTPWANSDYRLNRTPTQDQCT--KKYFTEGG 404
QY 417 DPNHMGGFDFLLANDVDLSNPVQAEQLNQHILYLMNWSIVMGDKDANFDGIRVDAVDNV 476
Db 405 E-----GGYEFLLSNDVDSNPVQAEQLNQLHLYLMNMGDIVMGDKDANFDGIRVDAVDNV 460
QY 477 DADMLQLYTNTYFREYGVNKSSEANALAHISVLEAWSLNDNHNDKTDGAALAMENKORLA 536
Db 461 NADLLQVTSNTPKONYKVTDSSEANALAHISVLEAWSLNDNHNDKTDGAALAMENKORLA 520
QY 537 LFLSIKAPIKERTPAVSPLYNNTFTTQDEKTDWINKDGSKAYNEDGTQKSTIGKYN 596
Db 521 SLAVLTK-----PQGRIDLNLISSEVNERAND-----T 551
QY 597 KYGDASGNYVFIKRAHDNNVDIIABIIKKEINPKSDGFTTDAEMKQAFIYNKMDLSSD 656
Db 552 AYGDPTIPTYSFVRAHDSVQTVIAKIVKEKIDTNSDGYTFTTLDQLKDAFKIYNEDMAKVN 611


```
QY 657 KYITLNNIPAAVAVMLQNMETITRVYVYGLYTDGHHMETKSPYVDTIVNLMSRIKYVS 716
Db 612 KYITHYNIIPAAVALLSNMESPRVYVYGLYTDGQYMAKSPYYDAIATMLQGRYAYVS 671
QY 717 GGOAORSYWLPTDGRKMDNSDVELYRTNEVYTVSRVYVYGLYTDGHHMETKSPYVDTIVNLMSRIKYVS 776
Db 672 GQOSEVH-----KNGNQLSSVRVYVYGLYTDGHHMETKSPYVDTIVNLMSRIKYVS 718
QY 777 ANPNKLNLQSAKLAVEMKIHANQKRYALIVGTADGINKFTSDADAIAGVYVYGLYTDGHHMETKSPYVDTIVNLMSRIKYVS 836
Db 719 SNDPNLDLGDSD-LTVNMGRAHQAQYRPLILGTGQVQSLKSD--TNIVKYTDANG 774
QY 837 VITFGANDIKGVETFDMSGFVAVVPGASDNQDTRVAPSTEAKKEGELTLKATEYDSQ 896
Db 775 NLTFADDIKGYSTVDMGYLAVVPGAKQDQVRVAADTNQKADGK-SLKTSAAALDSQ 833
QY 897 LIYEGFSNFQTTIPDGSDDPSVYTNKIAENVDLFKSWGVTSEFMAQFVSADGTFLDSVI 956
Db 834 VIYEGFSNFQDF--ANNDADYTNKIAENADFFKLGITTSFEMAPQYVSATDGSFLDSII 891
QY 957 QNGYAFADRYDLAMSKNNKYSGKEDLRDALKALHAGIOAIADWVDPQIYQLPGKEVYTA 1016
Db 892 QNGYAFSDRYDLAMSKNNKYSGKODLANALKALHAGIOAIADWVDPQIYQLPGKEVYTA 951
QY 1017 TRTDGAGRIKIAIDHLSYLVANSKSGDYQAKYGGFELAEIKAKYPMFKVMIMISTGK 1076
Db 952 KRTNSYGNPTFDAYINNALYATNTKSSGSDYQAGYGAFLDELKAKYPMFTVMIMISTGK 1011
QY 1077 PIDDSVKLQWKAEPNGTNVLERGVYVYGLYTDGHHMETKSPYVDTIVNLMSRIKYVS 1136
Db 1012 PIDPSTKIKQWKAEPNGTNVLERGVYVYGLYTDGHHMETKSPYVDTIVNLMSRIKYVS 1071
QY 1137 GFSSDGGKTYFTGTSQAKSAFVTFNGNTYTFDARGHMTVNTSEYSPNGKDVYRPLNGI 1196
Db 1072 GFYVDTGWAYYSTSGNAKAVSFYEGHYVYFDKDHMTVNTSEYSPNGKDVYRPLNGI 1130
QY 1197 MLSNAPYIDANGNTYLYNSKGOMYKGG--YTFPDSVSETDKGKSKVYFRFTNEGVMA 1254
Db 1131 QMRDAIYODAOGNSYYIGRTGILYKGDNWPYFVDPNNANK-----TVFRYFDANNVNA 1183
QY 1255 KGVTVIDGFTQYFGEDGFOAKDLVTFKGTYYFDDAHTGNGIKDWRNNGKWIYFDANG 1314
Db 1184 IGRNMYGOTYTFDENGFOAKQLLTDKGTHTYFDDGNGAMAKNFVNVGDDWYMDGNG 1243
QY 1315 VAATGAQVINGOKLYFN-EDGSQVKGVVYVKNADGTYSKYKEGFGELVTFNEFFTTDGNVWY 1373
Db 1244 NAVKGOYPVNNQILYFNPETGVQVKGQFITDAQRTSYDYDANGSALKSSGFFTPNGSDWY 1303
QY 1374 YA-----GANGKTVTGAQVINGOHLYFNAD-GSOVKGVVKN 1409
Db 1304 YAEYGVYVYKQVAENQDQWYFDDTTGKQAKGAQVDDGRDLVFNPDSPGVQKGFATD 1363
QY 1410 ADGTYSKYNASTGERLTNEFFTTGDNWYVYIGANGKSVTGEVKI-----1453
Db 1364 ESGNTSFYHGDNGDKVVGFFFTGNNWYVYADNNGNLVKGFEIDGKWTHTFDEVTCQQAQ 1423
QY 1454 -----GDDTYFFAKDGKQVKGQTSAGNGRISYYVYDGSKEAVSTWIBIQPGVYVYFDK 1507
Db 1424 GAALVNGQQLYFDVDSGIVQKGFVTDGQNTSYDVNSGKKVNGFFTTGDNWYVYADG 1483
QY 1508 NG 1509
Db 1484 QG 1485
```

RESULT 4

```
US-10-797-821-38
; Sequence 38, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
```

```
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 38
; LENGTH: 1554
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
; US-10-797-821-38
```

Query Match 48.2%; Score 3818; DB 5; Length 1554;

```
Best Local Similarity 50.1%; Pred. No. 1.2e-198;
Matches 783; Conservative 215; Mismatches 434; Indels 130; Gaps 24;
```

```
QY 1 MENKHYKHLKHKVQWVIAVASVALATVLGLSVTSSVSADETQDKTVTOSNSTTAS 60
Db 1 MEKHLHYKHLKHKVHWVIAVASIGLVSLV-----AGTVSA--EDKVANDTTAQTAVG 52
QY 61 LVTSPEATKEADKRNITKEADVLTTPAKETNAVETAT-TTNTQATAEAAATATTADVA-VA 118
Db 53 VDTGQDQATNTDANTNTTDTADQASANTNQDQAGSDQSNNDQQAQDQTANTDRNQADS 112
QY 119 AVPNKEAVVTTDAPAVTTEKABEQATVKAEVVNTEVKAPEALKDSEVAALSLKNIKN 178
Db 113 QPDNNOATDQATSPATDGTSVORRDA--ANVATAADQEGQTPAPSEKSAALSLDNVKL 169
QY 179 IDGKYVYVNEGSGSHKENFAITVNGQLLYFGKD-GALTSSSTYSFTPTGTTNIVDGFINN 237
Db 170 IDGKYVYVQADSGYKKNFAITVNGQLLYFDSDTGALSTSTYSFSQGTTLNLDVDFSSH 229
QY 238 AVDSSEASGELIDGYLTADSWTRPASIIKGVYVQASTAEADPRPLLMANVNDTQVNYL 297
Db 230 AYDSTAKSPELVNGVLTANSWYRPAIGLRNGQTWEASNENDLRPVLMSWPKDQTQVAYV 289
QY 298 NYMSKVFNL-DAKYSSTDQETLKYAAKDIQIKIEKIOAEKSTOWMLRETISAFVKTPQ 356
Db 290 NYMNKYL SANETEVNTS QVDLNKEAOSIQTKIBQITSDNSTOWMLRTAMEAFVAAQPK 349
QY 357 MNKETENYSGGGEHLOGGALLYVNDSTRTPWANSYRRLNRTATNTQGTIDKSIILDEOS 416
Db 350 WNNSTENFKG--DHLGGALLYTNSDLTPWANSYRLLNRTPTQODGT--KKYFTEGG 404
QY 417 DPNHGGGDFLLANDVLSNPVVAEQNLQIHYLNMWGSIVMGDKDANFDGRVDAVDNV 476
Db 405 E-----GGVEFLLSNDVNSNPVVAEQNLQIHYLNMWGSIVMGDKDANFDGRVDAVDNV 460
QY 477 DADMLQLYTNYPREYGVYVKNSEANALAHISLVLEAMSLNDNHYNDKTDGAALAMENKORLA 536
Db 461 NADLLQVYSNYPKDNKYKVTDSSEANALAHISILEAMSLNDNQNDNEDTNGTALSIDNSRLT 520
QY 537 LLFLSLAKPIKERTPAVSPLYNNTFNTRDEKTDWINKDGSKAYNDKGTQSTIGKYNE 596
Db 521 SLAVLTK-----PQGRIIDLNLISESVNKERAND-----T 551
QY 597 KYGDASGNVYFTRAHNNVQDIIAEIIEKKEINPKSDGFTITDAEMKQAFIYNKMLSSD 656
Db 552 AYGDITPTYSFVRAHDSQVTVIAKIVKEKIDTNSDGYTFTLDQLDAFKIYNEDMAKNV 611
QY 657 KKYTLNNIPAAVAVMLQNMETITRVYVYGLYTDGHHMETKSPYVDTIVNLMSRIKYVS 716
```



```
Db 613 LKQAFKYNEDMRQAKKKTQSNIPAYALMLSNKDSITRLYYGDMYSDGQYMATKSPY 672
QY 701 YDTIVNLMSRKIKYVSGQAQRSYWLPDQKMDNSDELRYRTNEVYTSVRYGKDIWAND 760
Db 673 YDAIDTLLKARIKYAAGGQDMKITVY--EGDKSHMD---WDYTVGLTSVRYGTGANEATD 727
QY 761 TEGSKYSRPSGGQVTLVANNPKLNDQSAKLNVEMGKIHANOKYRALLIVGTADGKFNPSD 820
Db 728 -QGSEATKTQGMVITSNPNLSKLNQNDKVIYVMGAAHKNQYRPLLLTKDGLTSYTS- 785
QY 821 ADAIAAGYVYKETSNGVLTFGANDIKGYETFDMSGFVAVMVPVGASDNODIRVAPSTEAK 880
Db 786 -DAAKSLYRKNDKGELVDFDASDIQGYLNPQVSGYLAVMVPVGASDNQDVRVAASNKAN 844
QY 881 KEGELTLKATEAYDSQLIYEGFSNFQT-IPDGSFSPVYTRNRKIAENVDLFSKSWGTSFPM 939
Db 845 ATGOV-YESSALSQLIYEGFSNFQDFVTKDSD---YTNKKIAQNVLFSKSWGTSFPM 900
QY 940 APOFVSADGTFLDSVIQNGYAPADRYDLAMSKNNKYGSKEDLRDALKALHAGIAQIAD 999
Db 901 APOYVSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSQDMINAVAKALHKSQIQUIAD 960
QY 1000 WYPDQIYLPGEVVTATRTDAGAKIADAIIDHSLYVANSKSGKDYOAKYGGEFABL 1059
Db 961 WYPDQIYLPGEVVTATRTVNDYGEYRKDSEIKNTLYAANTKSNKGDYOAKYGGAFLSL 1020
QY 1060 KAKYPMFKVNMISTGKPIDDSYKLGKQWKAEPNGTGNVLERGYYVGLSDSATGKYFTVTK 1119
Db 1021 AAKYPSIFNRTQISNGCKIDPSEKITAWKAKYFNGTNIILGRGVYVLDKNASDKYFELKG 1080
QY 1120 EGNFPLQUTGKEKVTGFSDDGKGYTFGTSGTQAKSAFV-TFNGNTYTFDARGHMVTN 1178
Db 1081 NOTYLPKQMTNKE-ASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYTFDNGHMVYG 1139
QY 1179 SEVSPNGKDYVFLPGLNGIMLSNIFYIDANGNTYLYNSKGMYKGYTKFDVSETDKDGE 1238
Db 1140 LQ-QLNG-EVQFSLNGVQLRESFLENADGSKNYFGLGNRYSNHYYSFDNDS----- 1190
QY 1239 SKVYFRYFTNEGVMKAGVTVTIDGFTQYFCEGDFQAKDKLVT-FKGKTYTFDAHTNGIK 1297
Db 1191 ----KWRYFDASGVMAVGLKTINGNTQYFDQDGYQVKGAWITGSDGKKRYFDGSGSNMAV 1246
QY 1298 DTRWN-INKWYTFDANGVAATGAQVINGOKLYFNEDGSOVKGVVKNADGTYSKYKEGF 1356
Db 1247 NRFANDKNGDWYLYNSDGIALGVQTINGKTYTFGQDGKQIKGIITD-NGKLKYFLANS 1305
QY 1357 GELVTNEFTTDCNVWYVYAGANGKTVTGAQVINGOHLVFNADGSOVKGGVKNADGTYSK 1416
Db 1306 GELARNIFATDSQNNWYTFSGDGAVTGSGTTAGKKLYFASDGKQVKGSEV-TYNGKVHY 1364
QY 1417 YNASTGERLTNEFTTGDNNWYVYIGANGKSVTGEVKIGDDTYFFAKDGKQVKGQTVSAGN 1476
Db 1365 YHADSGELQVNRFEADKNGWYLYDSNGEALTGSGQRINQDQVFFFTREGKQVKG-D-VAYDE 1423
QY 1477 GRISYY 1482
Db 1424 RBLVY 1429
```

RESULT 6

```
US-10-383-930-36
; Sequence 36, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
```

```
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 36
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-36
```

```
Query Match 44.2%; Score 3502.5; DB 4; Length 1430;
Best Local Similarity 48.1%; Pred. No. 1.5e-181;
Matches 724; Conservative 247; Mismatches 434; Indels 101; Gaps 32;
```

```
QY 1 MENKHYKHLKHKQWVTIAVASVALATVLGGLSVTTSSVSAD----ETQDKTVTQ-SNSG 56
Db 1 METKERYKHKYKHWVTVAVAS-GLITL--GTTTLGSSVSAETEQQTSKDVVTKSEDD 57
QY 57 TTASLAVTSPEATKKEADKRTNTKEADVLTPAKETNAVETATTTNTQATAEAAAT-ATTADV 115
Db 58 KAASSTQTDAPKTQQAQTEQTQAO-----SQANVADTSTSTKETPSONIITQANSDDK 112
QY 116 AVAAVPNKEAVVTTDAPAVTTEKAEQEPATVKAEEVNVNTEVKA-----PEAALKDSEVEA 169
Db 113 TVTNTKSEEAQTSSE-----RTKQSEEAQTASSQAL-TQAKAELTKQRTAAQENKNPVD 167
QY 170 ALSLKNINIDGKYVYVNEEDGSHKENFALTVNGQLLYFGKD-GALTSSSTYSFTPGTTNI 228
Db 168 LAAIPNVKQIDGKYIYIGSDGQPKNFALTVNNKVLDFKNTGALTDTDSQYQKQGLTKL 227
QY 229 VDGFSINNRAYDSSEASFELIDGYLTADSVYRPASI IKDGVTTQWASTAEDFRPLLMAMP 288
Db 228 NNDYTPHQIVNFENTSLETIDNYVTADSWYRPKDLKNGKWTWASSSDLPPLMSWPP 287
QY 289 NYDVTQVNLNYSKY-FNLDAKYSTDKQETLKVAAKDIQIKIEQIKQAEKSTQWLRETI 347
Db 288 DKQGTQIAYLNYNQOGLGTGENYTADSSQESLNLAAQTQVQVKIETKISQTOQTQWLRLDI 347
QY 348 SAFVTKTQPMKEBETNYSKGGEDHLOGGALLYVNDSTRTPWANSYRRLNRTATNQTGTI 407
Db 348 NSFVKTPQNMWNSQTSSTDSAGEKHLQGGALLYSNSDKTAYANSYRLLNRTPTSTQTK- 406
QY 408 DKSILDEQSDPNHMGDFDILLANDVLSNPVVOAEOQLNOIHYLMWNGSVTMGDKDANFQ 467
Db 407 -----PKYFEDNSSGGYDFLLANDIDNSNPVVOAEOQLNHLMYLNGSVLVANDPEANFQ 461
QY 468 IRVDAVDNVDADMLQLYTNFYREYGVNKSSEANALAHISVLEAWSLNDNHNDKTDGAAL 527
Db 462 VRVDAVDNVDNADLLQIASDYLKAHGYVDKSEKNAINHLSILEAWSNDNDPQYNKDTKGAQL 521
QY 528 AMENKORLALLPSLAKPIK-----ERTPAVSPLYNNTFTTQORDEKTDWINKGSKAY 580
Db 522 PIDNKLRLSLLYALTRPLEKDAKNKNEIRSGLEPVITNSLN----- 562
QY 581 NEDGTVKOSTICKYNEKYGDASGNVYFIRAHNNVDQIIABIIKKEINPKSDGFTITDAE 640
Db 563 -----NRSABEGKNSER-----MANTIFIRAHNSEVQTVJAKIIKQINPKTDLGLTFLDE 612
QY 641 MKQAFEINYKMDLSSDKKYLNNIPAAVAVMLQNNMETITRVYVYGDLYTDDGHYMETKSPY 700
Db 613 LKQAFKIYNEDMRQAKKKTQSNIPAYALMLSNKDSITRLYYGDMYSDGQYMATKSPY 672
QY 701 YDTIVNLMSRKIKYVSGQAQRSYWLPDQKMDNSDELRYRTNEVYTSVRYGKDIWAND 760
Db 673 YDAIDTLLKARIKYAAGGQDMKITVY--EGDKSHMD---WDYTVGLTSVRYGTGANEATD 727
QY 761 TEGSKYSRPSGGQVTLVANNPKLNDQSAKLNVEMGKIHANOKYRALLIVGTADGKFNPSD 820
Db 728 -QGSEATKTQGMVITSNPNLSKLNQNDKVIYVMGAAHKNQYRPLLLTKDGLTSYTS- 785
QY 821 ADAIAAGYVYKETSNGVLTFGANDIKGYETFDMSGFVAVMVPVGASDNODIRVAPSTEAK 880
Db 786 -DAAKSLYRKNDKGELVDFDASDIQGYLNPQVSGYLAVMVPVGASDNQDVRVAASNKAN 844
```



```
Db 901 APQYSSSEDSFLDSIIQNGYAFEDRYDLAMSKNNKYGSQODMINAVKALHKSIGQVIAD 960
Qy 1000 WVPDQIYQIPGKEVVTATFTDAGARRIADAIIDHSLYVANSKSKGYQAKYGGSEFLABL 1059
Db 961 WVPDQIYNLPKGKVVVATRVNDYGEYRKDSEIKNTLYAANTKSGNDKYQAKYGGAFSEL 1020
Qy 1060 KAKYEMKPMNISTGKPIDDSVKLQWKAERYFNGTNNVLERGYYVLSDEATGKYFTVTK 1119
Db 1021 AAKYPSIFNRTQISNGKKIDPSEKITAUKAKYFNGTNIILGRGVGYVLLKDNASDKYFELKG 1080
Qy 1120 EGNFIPLQLTGKEKVTGFSDDGKITTYFGTSGTQAKSAFV-TFNGNTYFFDARGHMYTN 1178
Db 1081 NOTYLPKQNTKE-ASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYFFDNGHMYVG 1139
Qy 1179 SEYSPNGKDVYRPLPNGIMLSNAFYIDANGNTYLYNSKQMYKGYTFKPDVSETDKDGKE 1238
Db 1140 LQ-QLNG-EVQYFSLNGVQLRESFLENADGSKNYFGHLGNYSNGYSPDNDS----- 1190
Qy 1239 SKVVRERYFTNGVMAGVTVIDGFTQYFEGEDGFOAKOKLVT-FKGKTYFFDAHTGNGIK 1297
Db 1191 ----KWRYFDASGVMAVGLKTNGNTQYFDQDGYQVKGAWITGSDGKKRYFDDGSGNMAV 1246
Qy 1298 DTRWN-INGKWTYFFDANGVAAATGAQVINGOKLYFNEDGSQVKGGVVGNADGTYSKYKEGF 1356
Db 1247 NRPANDKGDWYLYNSDGLALGVQTINGKTYFFGQDGQIKGIITD-NGKLKYFLANS 1305
Qy 1357 GELVTNEFTTDGNWYVYAGAKVTGQAVINGOHLYPNADGSQVKGGVVKNADGTYSK 1416
Db 1306 GELARNIFATSDQNNWYFSGDVAVTGQTIAGKKLYFASDGKQVKGsfv-TYNGKVHY 1364
Qy 1417 YNASTGERLTNEFTTGDNNWYVYIGANGSKSVTGEVKIGDDTYFFAKDGKQVKGQTVSAGN 1476
Db 1365 YHADSGELQVNRFEADKQGNWYLYDSNGEALTGQSQRINDQRVFFTREBQVKGd-VAYDE 1423
Qy 1477 GRISY 1482
Db 1424 RLLVY 1429

RESULT 8
US-10-383-930-37
; Sequence 37, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 37
; LENGTH: 1590
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-383-930-37

Query Match 41.9%; Score 3325.5; DB 4; Length 1590;
Best Local Similarity 45.0%; Pred. No. 7.2e-172;
Matches 710; Conservative 239; Mismatches 456; Indels 173; Gaps 33;

Qy 1 MENKHYKLHKYKKQWVTIYASVA-LATVIGLSVTTSSVGADETDQKTVTQNSGTTA 59
Db 1 MEKNVRFKWHKVRWVTVLSVASATMLASALG-----ASVASAD-----TDTASDD----- 46
Qy 60 SLVTSPEATKEADKRNTEADVLTPAKETNAVETATTTTQATABAAATATTADVAVAA 119
Db 47 ----SNQAVVVGDTNNQATD-----QTSNAATAT-----SEQSASTDAATDQASAA 90
```

```
Qy 120 VPKNEAVVTTDAPVTTTEKAEQEPATVKAEVVN-----TEVKAPEAALKDSEVEAL 171
Db 91 EOTOGETTASTTAAQTQTTTNAME-AKWVPTENENOGFTDEMLAEAKNVATABSDSIPSDIA 149
Qy 172 SLKNIKNDGKYVYVNEGSHKENFAITVNGQLLYFGKDGALTSSTSYSTFTGTTNIVDG 231
Db 150 KMSNVKQVDGKYVYDQDGNVKKFPAVSGDKIYYFDETGAYKDTSKVDADKSSSAVSQN 209
Qy 232 ---PSINRKYVDSSEASPELIDGYITADSWTRPASIIKDGVTWQASTADPRPLMAWVP 288
Db 210 ATIFAANNRAYSTSAKNFEAVDNYLTADSWTRPKSILKDGKTWTESGDKDPRPLMAWVP 269
Qy 289 NVDQVNYLYNYSKVFNLDAKYSSYTDKQETLKVAAKDIOIKIEQIKQAEKSTQWLRETS 348
Db 270 DTETKRNVTYNNKVVGIDKTYTAETSQADLTAAAEALVQARIEQIKTSENNTKWLREALS 329
Qy 349 AFVKTQPOWNETEN-YSGKGGEHLQGGALLYVNDV-RTPWANSDYRRLNRTATNQTGT 406
Db 330 AFVKTQPOWNESEKPY-----DHLQNGALLFDNQDTLTPDTQSNYRLNRTPTNQTS 384
Qy 407 IDKSILDSQSDPNHMGDFFLANDVLSNPVQAEQLNQIHYLMNWSIYVNGDKDANPD 466
Db 385 LDSRPTYNPNDP--LGGYDFLLANDVDNSNPVQAEQLNWLHYLLNFGSIYANDADANPD 442
Qy 467 GIRUDAVDNDADMLQLYTNYPREYGYVKNSEANLAHI SVLEAKSLNDNHNDKTDGAA 526
Db 443 SIRUDAVDNDADLLQISSDYLKAAAYIDKNNKANNHVSIVVEASNDNDTPYLHDDGDNL 502
Qy 527 LAMENKQRLALLFSLAKPIKERTPAVSPLYNNTFNTQDEKTDWINKDGSKAYNEDGTV 586
Db 503 MNMDKFLSLMWSLAKPLDKRS-CLNPLIHSNLDVREVDNR-----EVEIV 548
Qy 587 KQSTIGKNEYKYGDSAGNYVFIRAHDNNVQDIIABIIKKEINPKSDGFTITDAEMKQAPE 646
Db 549 P-----SYSPARAHDSVQDIIRDIKAEINPNPSFGYSFTQEEIEQAFK 592
Qy 647 IYNKMLSDSKKYTLNNTPAAYAVMLQNMETITRYVYGDLYTDDGHYMETKSPYDITVN 706
Db 593 IYNEDLKKTKKTYTHYNPLSYTLLTNKGSIPRVYVYGDYFDDGQYMWANKTVNYDAIES 652
Qy 707 LMKSRIKVSGGOAORSYWLPTDGMNDSDVLYRTNEVYTSVRYGKDOIIMTANDTEGSKY 766
Db 653 LLKARKKVVSGGOAMQNYQI-----NGEILTSVRYGKGAALKQSD-KGDAT 697
Qy 767 SRTSQVTLVANNPKLNDQSAKLVEMGKTHANOKYRALIVGTADGINKFTSSDADAIAA 826
Db 698 TRTSGVGVVMGNQPNFSLDGKV-VALNMGAAHANQEYRALMVSTKDG VATYATDADASKA 756
Qy 827 GYVKETSDNGVLTGANDIKGYETFDMSGFVAVVVPVUGASDNQDIRVAPSTAKKEGELT 886
Db 757 GLVKETDENGYLFLINDLKGVANPQVSGFLQVWPVPGAADQDIRVAASDASTASTDGK-S 815
Qy 887 LKATEAYDSQLIYEGFSNFQTIIPDGSPPSVYNNRKAENVDLFKSWGVTSSFEMAPQFVSA 946
Db 816 LHQDAAMDNRVMFEGFSNFQSP--ATKDEEYTNVVIANNVDKFSWGIITDFEMAPQVSS 873
Qy 947 DDGTFELDSVQNGYAPADRYDLAMSKNNKYSGKEDRLDALKALHAKAGIOAIADWPDQY 1006
Db 874 TDGQFLDSVQNGYAFTRDYDLGMSKANKYGTADQVKAIKALHAKGLVMADWVPDQMY 933
Qy 1007 QLPKGKVVVATFTDAGARRIADAIIDHSLYVANSKSKGYQAKYGGSEFLABLKAKYPEM 1066
Db 934 TFPKQEVVTVTRTKDFGPKPIAGSQNHSLYVDTTKSSGDDYQAKYGGAFDLBELKYPEL 993
Qy 1067 PKVNMISTGKPIDDSVKLQWKAERYFNGTNNVLERGYYVLSDEATGKYFTVTKEGNFIP 1126
Db 994 FTKKQISTGOALDPSVKIKQWSAKYFNGSNILGRADYVLSQVSKNKNYFNASDITFLPS 1053
Qy 1127 QLTGKEKVTGSSDKGITY-FGTSGTQAKSAFVTFNGNTYFFDARGHMYTNSEYSPNG 1185
Db 1054 SLLGK-VVESGIRYDGKGYIYNSSATGDQVKASFITEAGNLVYFGKDGVMYVGAQ-TING 1111
```



```
Db 1112 AN-YPFLENGTALRNTIYTDAGNSHYANDCKRYENGYYQF-----GND-----WR 1157
Qy 1246 YFTNEGVMAGVTVDIGFTQYFEGEDGFOAKOK-LVTFKGKTYFPAHTGNGIKDTW-RNI 1303
Db 1158 YF-KDGNMAVGLTVDGNYQYFDDKGQVQAKDIIVTRDGKRYFDPQHNGNAATNTFIADK 1216
Qy 1304 NGKWTYFDFGAAVGAQVINGOKLYFNEGSGOVKGKVNADGTYVSKYKEGFGELVTNE 1363
Db 1217 TGHWYLGKDGAVTGAQTVGQKLYFEANGQQVKGDFVTSDEGKLYFYVDVSGDMWTD 1276
Qy 1364 FTTDGNVWYAGAGKTYTGQVINGQHLVFNADGSOVKGKVNADGTYVSKYNASTGE 1423
Db 1277 FIEDRAGNWFYLGKDGAAVTGAQTVGQKLYFEANGQQVKGDI VGTGDKIRYYDAKSGE 1336
Qy 1424 RUTNE-----PFTTG 1433
Db 1337 QVFNKTVKAADGKTYVIGNDGVAVDPVVKGTOTFKDASGALRFYNLKGQLVTGSGWYETA 1396
Qy 1434 DNNWYIYGANGSVTGEVKGIDDTYFFAKDGQVKGOTVSAGNGRISYYYGDSGKRAVST 1493
Db 1397 NHDWYI-OSGKALTGBQTNGHLYFKEDGHQVKGQLVTGDKVRYDYDANSQQAFNK 1455
Qy 1494 WIEIOPGVVYVFDKNGLA 1511
Db 1456 SVTVNGKTY-YFGNDGTA 1472

RESULT 10
US-09-740-274-2
; Sequence 2, Application US/09740274
; Patent No. US20020031826N1
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-740-274-2

Query Match 40.3%; Score 3196.5; DB 3; Length 1475;
Best Local Similarity 45.2%; Pred. No. 6.7e-165;
Matches 695; Conservative 233; Mismatches 472; Indels 139; Gaps 31;

Qy 1 MENKIHKLHKVKKOWVTIYAVS--VALATVLGGLSVTTSSVSADTQDKTQVTSNGT 58
Db 1 MDKKVRYKLKRVKRVTVSVASAVMTLTLSGL-----VKADSNESKQISNDNNTS 54
Qy 59 ASLVTSPKATKADKKTNTKEADVLTPAKETNAVETATTNTQATBAATATTADVAVA 118
Db 55 -----VVTANESNVITEATSKQEAASQTNHTVTVTSSTSS 91
Qy 119 AVPNKEAV---VTTDPATVTEKABEQPATV---KAEVVNTEVKAPEA---ALKDS--- 165
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 20:42:47 ; Search time 8.6381 Seconds
(without alignments)
2306.008 Million cell updates/sec

Title: US-10-797-821-40

Perfect score: 7928

Sequence: 1 MENIHYKLUHKVQWTTIA.....PGVYVYFDKNGLAYPRVLN 1518

Scoring table:

BLOSUM62

Gapbp 10.0 , Gapext 0.5

Searched: 97014 seqs, 13122538 residues

Total number of hits satisfying chosen parameters: 97014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA_New.*

- 1: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB_PUB.pep.*
- 2: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubppaa/PCT_NEW_PUB_PUB.pep.*
- 5: /cgn2_6/ptodata/1/pubppaa/US05_NEW_PUB_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubppaa/US10_NEW_PUB_PUB.pep.*
- 7: /cgn2_6/ptodata/1/pubppaa/US11_NEW_PUB_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB_PUB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	522	6.6	2710	7	US-11-051-453-41
2	367	4.6	2367	7	US-11-051-453-42
3	338.5	4.3	1992	7	US-11-013-759-3
4	338.5	4.3	1992	7	US-11-013-759-13
5	338.5	4.3	2047	7	US-11-013-759-4
6	338.5	4.3	2047	7	US-11-013-759-7
7	322	4.1	396	7	US-11-022-562-228
8	316.5	4.0	2053	7	US-11-013-759-9
9	301	3.8	726	7	US-11-013-759-11
10	299	3.8	2314	7	US-11-013-759-11
11	280.5	3.5	619	7	US-11-052-554A-229
12	276.5	3.5	701	7	US-11-052-554A-231
13	267	3.4	744	6	US-10-873-528-184
14	255.5	3.2	1647	7	US-11-052-554A-260
15	251	3.2	693	6	US-10-873-528-185
16	246.5	3.1	1767	7	US-11-052-554A-372
17	244.5	3.1	5024	6	US-10-793-626-2964
18	236	3.0	1155	6	US-10-793-626-1780
19	232.5	2.9	5291	7	US-11-052-554A-281
20	226	2.9	627	6	US-10-873-528-191
21	225	2.8	2902	7	US-11-052-554A-91
22	224	2.8	332	6	US-10-873-528-190
23	224	2.8	1588	7	US-11-052-554A-280
24	221.5	2.8	1263	6	US-10-485-517-127
25	220	2.8	690	7	US-11-052-554A-232

ALIGNMENTS

RESULT 1

US-11-051-453-41

; Sequence 41, Application US/11051453

; Publication No. US20050287150A1

; GENERAL INFORMATION:

; APPLICANT: AMBROSINO, DONNA

; APPLICANT: BABCOCK, GREGORY J.

; APPLICANT: BROERING, THERESA

; APPLICANT: GRAZIANO, ROBERT

; APPLICANT: HERNANDEZ, HECTOR JAVIER

; APPLICANT: LOWY, ISRAEL

; APPLICANT: MANDELL, ROBERT

; APPLICANT: MOLRINE, DEBORAH

; APPLICANT: THOMAS, JR., WILLIAM D.

; APPLICANT: ZHANG, HUI-PEN

; TITLE OF INVENTION: ANTIBODIES AGAINST CLOSTRIDIUM DIFFICILE TOXINS AND

; FILE REFERENCE: MJ1-001

; CURRENT APPLICATION NUMBER: US/11/051,453

; PRIOR FILING DATE: 2005-02-04

; PRIOR FILING DATE: 2004-02-06

; PRIOR APPLICATION NUMBER: 60/542,357

; PRIOR FILING DATE: 2004-09-28

; NUMBER OF SEQ ID NOS: 82

; SOFTWARE: PatentIn Ver. 3.3

; SEQ ID NO 41

; LENGTH: 2710

; TYPE: PRT

; ORGANISM: Clostridium difficile

US-11-051-453-41

Query Match 6.6%; Score 522; DB 7; Length 2710;

Best Local Similarity 20.1%; Pred. No. 2.2e-20;

Matches 318; Conservative 194; Mismatches 491; Indels 576; Gaps 71;

QY 176 IKNDGKYYVVEDGSHKENFA--ITVNGQLYFGKDGALTSSTVSFTPTGTTNIVDGRS 233

1438 ISNLSNTEKINTGLDLSKNIAVNTYDSENNKYFGAISKTSQKSIHYKKDSKNILFEYN 1497

QY 234 INNRAYDSEASFELIDGYLTADSWYRPASIIKDGVTWOASTAEDFRPLLMAMPNVDQ 293

1498 DSTLFSNKFDAEDINVFMKD-----1520

QY 294 VNYLNMKVFNLDAKYSTDQETLKVAAKDIOIKIEQIOAEKSTQWLRETISA----349

1521 ---INTITGKYVD---NNTDK-----SIDFSISLVSKNQVKVNGLYLNESVSYLD 1567

Db 745 IADQSRNIEL-----GNTIQDKSNAASIND-----ILNTGFLKNNNNPI 787
QY 298 NYMSKVFNLD- AKYSTDKOETLKVAAKDQI KIEBQIOAKSQTOWLRRETI SAFVKTOQP 356
Db 788 DFVSTYDVIDFANGNATTATVTHDTANK--TSKVVDVNVDDTTIHLT-----833
QY 357 WNKETENYSGGGEHLGGALLVYNDSTPWNSDYRELNTATNTQTGIDKSILDEQS 416
Db 834 -----GTDDNKKLGV-----KTTKLKNTSANGNTATNPNVSSDE 868
QY 417 DPNHMGFFDLANDVLSNPVVQABOLN-----QIHLYLMNWSI VMGDKDANFDGIRVDA 472
Db 869 D-----ALVNAKDI-----AENLNTLAKEIH-----TTKGTADTALQIFTVYK 906
QY 473 VD-NVDADMLOLYTNYFREYYGVNKSSEANALAHISVLEAWSLNDHNDKTDGAALAMEN 531
Db 907 VDENNANADANAIT-----VGQKANNQVNTLTAK-----GENGLNIKTD-----946
QY 532 KQRLALLPSLAKPIKERTPAVSPLYNNTTQORDEKTDWINKDGS KAYNEDGTVKQSTI 591
Db 947 -KNGVTVTFGI-----NTTSGLKAGKST--LNDGGLSIKNPTGS-EQIQV 986
QY 592 GKYNKYGDASGNVYFIRAHNNVDIIAEIILKEINPKSDGFTITDAEMKQAFELYNKD 651
Db 987 GADGVKFAKVNNGVVGAGIDGTR-----ITRDEI-----GFTGTGSLDKSKHLSKD 1036
QY 652 MLSSDKKXTLNNIPAAVYMLQNMETITRVYVYGDLYDDGHYMETKSPYYD---TIVNLM 708
Db 1037 GINAGGK-----KITNIQSGEI-AQNSHDAVTGGKTYDLKTELENKI 1077
QY 709 KSRIVSGGQORSWLPFTDGMKMSDVLELRTHEVYTSVRYGKDIMTANDEGSKYR 768
Db 1078 SSTAK-----TAQNS-----LHEFSVADQGNFTVSNPYSSYDTSK 1114
QY 769 TSGOVLVANN-----PKLNLDQSACL-----NVEMKIHANOKYALIVG 809
Db 1115 TSDVITFAGENGITTKVNGVVRVIGIDQTKGLTTPKLTVCNNNGKIVIDSQNGQNTING 1174
QY 810 TADGINKFTSDADAI-----AAGVYKETS DN-----GVLTPGAN-----DIKGYET 850
Db 1175 LSNLTIANVTNDKGSVRTTEQGNIIKDEDKTRAASIVDLSAGFNLGNGEAVDFVSTYDT 1234
QY 851 FDMSGFVAVVVPVGSADNQ-----DIRVAESTEAKKEGEL-----TLKATEAVDSQLI 898
Db 1235 VNPADGNATTAKVYDTSKTSKVVDVNVDDTTIEVDKDLGVKTKTTLTSTGTGANKPA 1294
QY 899 YEGFSNFQIPGSDPSVYTRKIAENVDLFKSWGVTSPFEMAPQFVSADDTGLDSVIGN 958
Db 1295 LSNQATGDALVKASDIVAHLN---TILSGDIQTAKGASQANNAGYVDADGNKVI-----1345
QY 959 GYAFADRYDLAMSKNNKY--GSKEDLRDALKALHAGIOAIADWPDDQIYQLPGKEVVA 1016
Db 1346 -----YD---STDNKYYQAKNDGTVDTKTEVAKDKLVAQAQTPDGTGLAQNMVKSVIN 1395
QY 1017 TRTDGAGRKIADAIIDHSLYVANSKSSGKYQAKYGGEPFLAEK--AKYPEFKVNMIST 1074
Db 1396 EQVNDANKK--QGINEDNAPVKGLEKAASDNKTKNAAVTVGDNLNAVAQTPFLTFAGDTGTT 1453
QY 1075 GKPIDDSVKLKQWKAERYFNGTIVL-ERGVGYVLSDEATGKYFTVTKEGNFIPQLTGKEK 1133
Db 1454 AKKLGETLTKGGQTD---TNKLTNNIGVVAGTDG---FTVK-----LAKDLTNLNS 1500
QY 1134 VITGSSS-DKGITYFTGTS--TOAKSAFVTFNGNTYVYFARGHMVNTNSYSPNGKDVYRF 1191
Db 1501 VNAGGTKIDDKGVSFVDSGQAKANTPVLSSANG---LDLGGKVISNVGKGTOKTDAAV 1556
QY 1192 -----LPGNIMLSNAPFIIDANGNTYLYNSKGQYKGYTKFPDVSETDKDGSKSVKVPY 1246
Db 1557 QQLNEVRNLLGLGNAGDNADGN-----QVNIADIKKDPNNGS-----1594
QY 1247 FTNEGVMAGVTVIDG---FTQYFGEDGFQ-AKDKLVTFKGTYYVFAHTGNGIKDWT- 1300
Db 1595 -SSNRVIVKAGTVJGKGNNDEKLATGIGVGVK-----DGNANGDLSNVWV 1642

RESULT 4

US-11-013-759-13
; Sequence 13, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loomore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN OF MORAXELLA
; FILE REFERENCE: 1038-921MIS:jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 1992
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-13

Query Match 4.3%; Score 338.5; DB 7; Length 1992;
Best Local Similarity 19.5%; Pred. No. 1.4e-10;
Matches 315; Conservative 228; Mismatches 633; Indels 437; Gaps 73;

QY 33 LSVTTSSVSADETQDKVTVTQSNSTGTASLVTSF-----EATKEAD-----KRTNT 77
Db 534 IGVKTTLSNDSGTSK-FSVKSGTNNLSILVTAELHASVLYNEVNRVTSALQSFTVKEEDD 592
QY 78 KEADVLTTPAKETNAVETATTTNTQATAEAATATTADVAVAAPNKE--AVVTTDAPAVT 135
Db 593 DDANAITVAKD-----TTKNAGAVSILKLGKNGLTVATKKDGTVT 633
QY 136 TEKAEQEPATVKAEEVNVTEVKAPEAALKDSEVEAALSLEKNIK--NIDGKYVYVNEEDGSHK 193
Db 634 FGLSDSGSLTGKSTLND---GLTVKDTNEQIOVGANGIKFTNVG-----SNPGTGI 684
QY 194 ENFALTVMGQLLYFKDQALTSSTSY-----SFTPGTTNIVD-GFSINNRAY-----239
Db 685 ANTARITRDKIGCFAGSDGAVDTNKPYLDQDKLVGNVKTNTGINAGKAITGLSPTLPS 744
QY 240 --DSSEASFELIDGYLTADSVTRPASIIKDGVTWQASTAEDPRPLLMWPNVDTQVNYL 297
Db 745 IADQSRNIEL-----GNTIQDKSNAASIND-----ILNTGFLKNNNNPI 787
QY 298 NYMSKVFNLD-AKYSTDKOETLKVAAKDQI KIEBQIOAKSQTOWLRRETI SAFVKTOQP 356
Db 788 DFVSTYDVIDFANGNATTATVTHDTANK--TSKVVDVNVDDTTIHLT-----833
QY 357 WNKETENYSGGGEHLGGALLVYNDSTPWNSDYRELNTATNTQTGIDKSILDEQS 416
Db 834 -----GTDDNKKLGV-----KTTKLKNTSANGNTATNPNVSSDE 868

Qy	417	DPNHWGGDFLLANDVDLSNFPVQAEQLN-----QIHVLMNWGSIWMDKOANPFGIRVDA	472
Db	869	D-----ALVNAKOT-----AENLNTLAKSEIH-----TTKGTAOTFALQFTFTVKK	906
Qy	473	VD-NVDADMLOLYTNYFREYYGVNKSSEANALAHITSVLEAMSLNDHNHYNDKTGDGAALAMEN	531
Db	907	VDENNADDANAIT-----VGQKANNQVNTLTJK-----GENGLNIKTD-----	946
Qy	532	KQRLALLFSLAKPIKERTPAVSPLYNNFTNTQORDEKTDWINKDGSKAYNEDGVTKOSTI	591
Db	947	-KNGTVPVFGI-----NTTSLKAGKST--LNDGGLSINKNPTGS-EQIQV	986
Qy	592	GKYNEKYGDASGNVYFIRAHNNVQDIIAEIIKKEINPKSDGFTITDAEMKQAFEIYKND	651
Db	987	GADGVKFAKVNNGVVGAGIDGTTT-----ITRDEI-----GFTGTNGSLDKSKPHUSKD	1036
Qy	652	MLSSDKKYTLNPIPAAYAVMLQNNETITRVVYGLYTDGGHYMETKSPYYD---TIVNLM	708
Db	1037	GINAGGK-----KITNIQSGEI-AQNSHDAVTGCKIYDLKTELENKI	1077
Qy	709	KSRIKYVSGGOQRSYMLPTDCKMDNSDELRYRNEVYTSVRYGKOTMTANDTEGSKYSR	768
Db	1078	SSTAK-----TAQNS-----LHEFSVADEQNNFTVSNPSSYDTSK	1114
Qy	769	TSGOVTLVANN-----PKNLNDSAKI-----NVEMGKIHANOKYRALIIVG	809
Db	1115	TSDVITFAGENGITTKVNGVVRVGIQDTKGLTTPKLTGVNNGNGKGIVIDSQNGQNTITG	1174
Qy	810	TADGIKNFTSDADAI-----AAGYVKETDSN-----GVLTFGAN-----DIKGYET	850
Db	1175	LSNLTANVTNDKGSVRTTEQGNIIKDEKTRAASIVDLSAGFNLCNGEAVDFVSYDIT	1234
Qy	851	PDMSGFVAVVVPVCASNQ-----DIRVAPSTEAKKEGEL-----TLKATRAYDSOLI	898
Db	1235	VNFADGNATTAKVYTDYDTSKTSKVYVQVNVDDTTIEVKDKKLGKVTTLTSTGTGANKFA	1294
Qy	899	YEGFSNFQTTIPDGSPPSVYTRKIAENVDLFKSMGVTSFEMAPQFVSADDTGFLDSVIQN	958
Db	1295	LSNQATGDALVKASDIIVAHLN-----TLSGDIQTAKGASQANNAGIVDADGNKVI-----	1345
Qy	959	GYAPADRYDILAMSKNKKY--GSKEDLDLALKHAGIOAIADVPDQIYQLPCKEYVYTA	1016
Db	1346	-----YD-----STDNKKYQAKNDGTVDKTKEVAOKDLVAQAQTPDGTGLAQMNVKSIVNK	1395
Qy	1017	TRTDGCRKTAADAIIDHSLVYANSKSGKDYQAKYGGEFLAEUK--AKYPEMPKVNKMIST	1074
Db	1396	EQVNDANK--QGINEONAFVKGLEKAASDNKTGNAAVTYVGDNLNAVAQTFLTFAGDTGTT	1453
Qy	1075	GKPIDDSVKLQWKAERYFNGTNNVL-ERGVGVLSDEATGKYFTVTKEGNFIPQLQTCKEK	1133
Db	1454	AKLGETLITIKGGTD---TNKLTDNNGVAVGTDG---FTVK-----LAKDLTLNLS	1500
Qy	1134	VITGFSS-DGKGITYFTGSG-TQAKSAFVTFNGNTYYFDARGHVMNTSEYSPNGKDYRF	1191
Db	1501	VNAGGTKIDKGVSFVDSGQAKANTPVLSSANG---LDLGGKVISNVGKGTDXDTDAANV	1556
Qy	1192	-----LPGNMLSNVAFIDANGTYLYNSKGOMYKGGYTFKDFVSETDKOGKSEKVKFRY	1246
Db	1557	QQLNEVNTLGLGNAGNDADGN-----QVNTIADIKOPNSG-----	1594
Qy	1247	FTNEGVMVAKGVTVIDG---FTQYFGBDGFQ-AKDKLVTFKGTYYFDAGHTGNGIKDWT-	1300
Db	1595	-SSNPTVIKAGTVLGGKGNNDTEKLATGGIQQGVDK-----DGNAGDLSNVVV	1642
Qy	1301	-----RNINGKWYFDPANGVAATGAQVINGOKLYFNEDGSO---VKG--GV	1341
Db	1643	KTOKDGSKALLATYNAAGQTNVLTNNPAAIDRINEQGRFFRHVNDGNEPVPVQGRNGI	1702
Qy	1342	VKNADGTYSKYKGBFGLVNTNEFF-----TTDGNWVYVYAGANGKTVTGAQ-----V	1387
Db	1703	DSSAGSGHS-VAITGFQAKADGEAAVAIGROTQAGNOSIATGDNQAQ-ATGQDSIAITGNV	1760

Qy	1388	INGOHLYFNADGGSOVGGVKNADGTVS--KYNASGERLTNEF-----FTTGDNWYYI	1440
Db	1761	VAGHSGAIGDPSIVK-----ADNSYGVNNQFFDTQTDFGVGNNTVTVESVAL	1814
Qy	1441	GANGKSVTGEVKIGDDTFFPAKDGKQVKQGVTSAG-NGRISYYVYDSDGKRAVS	1492
Db	1815	GSN-----SAISAGTHAGTQAKSKSDGTAGTTTATGATGTGVKGFAGOTAVGAVS	1862

RESULT 5

```

US-11-013-759-4
; Sequence 4, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN OF MORAXELLA
; FILE REFERENCE: 1038-921MIS:jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2047
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-4

```

Query Match	4.3%	Score 338.5	DB 7	Length 2047
Best Local Similarity	19.5%	Pred. No. 1.4e-10		
Matches 315	Conservative 228	Mismatches 633	Indels 437	Gaps 73

Qy	33	LSVTTSSVSADETDQKTVTQSGNSCTTASLVTSP-----EATKEAD-----KRTNT	77
Db	589	IGVKTTELNSDGTSDK- FSVKGSGTNSLVTAEHLASVLYNEVNRDTASALQSFVKEED	647
Qy	78	KEADVLTPEAKETNAVETATTTNTQATAEAAATTATTADVAAPNKE--AVVTTDAPAVT	135
Db	648	DDANALTVAKD-----TTKNAGAVSILKUGKNGLTVATKDGCTVT	698
Qy	136	TEKAEQPAVTKAEVVNTEVKAPEAAALKDSEVAALSKNIK--NIDGKYVYVNEGDSHK	193
Db	689	FGLSQDSGLTIGKSLNND---GLTVKOTNEQIQVGANGICKETVNG-----SNPGTGI	739
Qy	194	ENFAITVNGQLLYFGKDGALTSSTY-----SFPFGTNNIVD-GFSINNRAY-----	239
Db	740	ANTARITRDKIGPAGSDGAVDTNKPDLQDKLVGNVKITNTGINAGGKAITGLSPTLPS	799
Qy	240	--DSSASPELIDGYLTADSWTRPASIIKDGVTWQASTAEDFRPLLMAWPNVDQTQNYL	297
Db	800	IADQSSRNIEL-----GNTIQDKKSNAASIND---ILNTGFNLKNNNP	842
Qy	298	NYMSKFNFLD-AKYSSTDQKTELKVAAKDIQIKIEQIOAEKSTQMLRETISAFVKTQPO	356
Db	843	DFVSTYDVIDFANGNATATVTHTDANK--TSKVYVDVNVDDTTIHL-----	898
Qy	357	WNKETENYKGGEDHLOGGALLYVNDSTRTPWANSDYRLNRTATNTQGTIDKSLDEQS	416
Db	889	-----GTDDNKKLGV-----KTYKLKNTSANGNTATNPNVSSDE	923
Qy	417	DPNHMGDFPLANDVLSNPVVQAEQLN----QIHLYLMNWSIVMGDKDANPDGIRVDA	472
Db	924	D-----ALVNAKDI-----AENLNTLAKEIH-----TTKGADTALQTFTVVK	961
Qy	473	VD--NVDADMQLQYTNVFRBYGYGNKSEANALAHISVLEAWSLNDNHYNKTGDAALAMEN	531
Db	962	VDENNADDANAIT-----VGQKNANQVNTLTAK-----GENGINTKTD-----	1001


```
1092 DB GINAGGK-----KITNIQSGEI-AONSHDAVTGKIYDLKTELENKI 1132
709 QY KSRIKVFSGQAORSYWLPTDGRWMSDVLYRTNEVTSVRGDKIMTANDTEGSKYSR 768
1133 DB TAQNS-----LHEFSVADEGNNFTVSNVSSYDTSK 1169
769 QY TSQVTLVANN-----PKNLIDQSAKL-----NVEMGKIHANQYRALIVG 809
1170 DB TSDVITFAGENGITTKVNGVVRVIGDQTKGLTTPKLTGVNNGKGIIVDSQNGQNTTIG 1229
810 QY TADGIKNFTSDADAI-----AAGVVKETDSN-----GVLTFFGAN-----DIKGYET 850
1230 DB LSNLTANVTNDKGSVRTTEQGNIIKDEKTRAASIVDVLVSAGFNLOGNGEAVDFVSTYDT 1289
851 QY FOMSGPVAVVWVPVGSADNQ-----DIRVAPSTEAKKEGEL-----TLKATEAYDSQLI 898
1290 DB VNPADGNATKATYDDTSEKTSKVVDVNVDDTTIEVKOKLGVKVTITSTGTGANKFA 1349
899 QY YEGFSNFOTIPDGSFVSVTNRKIAENVDLFKSWGVTSFEMAPQFVSADDTGTFDLSVIQN 958
1350 DB LSNQATGDALVRASDIVAHLN---TISGDIQTAKGASQANNSAGYVDADGNKVI-----1400
959 QY GYAFADRYDLAMSKNNKY---GSKEDLRDALKALHKAGIOAIAADWPDPQIYQLPGEVVTVA 1016
1401 DB -----YD---STDNKKYQAKNDGTVDKTEKVAQKLVQAQTPDGTGLAQMNVKSVINK 1450
1017 QY TRTDGAGRIADAIIDHSLYVANSKSSGKDYOQAKYGEFLAELK--AKYPEKFNVMIST 1074
1451 DB EQVNDANKK--QGINEDNAPVKGLEKAASDNKTQNAAVTVGDINAQAOTPLTFAGDTGTT 1508
1075 QY GKPIDDSVKLQWKABEYFNGTNVL--ERGVGYVLSDEATGKFTVTKEGNFIPQLTGTGKEK 1133
1509 DB AKKLGELTIKGGQTD---TNKLTNNIGVAGTDG---FTVK-----LAKDLTNLNS 1555
1134 QY VITGPSS--DGKITYFTGSG--TOAKSAFTVFNNGTNYTFDARGHMTVNSYSPNGKDVYRP 1191
1556 DB VNAGGTKIDDKGVSVFDDSSGQAKANTFVLSANG---LDLGGKVISNVGKGTGKTDAAV 1611
1192 QY -----LPGIMLSNAPYIDANGNTLYNSKGQMYKGYTKFDVSETDKDGKESKVVKFPY 1246
1612 DB QQLNEVRNLGLIGNAGDNADGN-----QVNIADIKKDPNSGS-----1649
1247 QY FTNEGVMAGKVTVIDG---FTQYFGEDEGQ--AKDKLVTFKGTYYFDDAHTGNGIKDTW- 1300
1650 DB -SSNRTVIKAGTVLGGKNNDEKLTATGIGQGVK-----DGNANGDLSNVWV 1697
1301 QY -----RNINGKYYFFDANGVAATGAQVINGOKLYFNEBDSQ---VKG--GV 1341
1698 DB KTQKDGSKKALLATYNAAGQTNVLTNNPAEIDRINEQGIREFHVDNGNQEPVVGQRNGI 1757
1342 QY VKNADGTSYKKEGEGELVTNEFF-----TTDGNVWYAGANGKTVTGAQ-----V 1387
1758 DB DSSASGKHS--VAIGFQAKADGAAVAIGRQTAQNSIAIGDNAQ--ATGDQSIATGTGNV 1815
1388 QY INGQHLVFNADGSQVKGKGVVKNADGYS--KYNASTGERLTNEF-----FTTGNNWYI 1440
1816 DB VAKHSIGAIGDPSVK-----ADNSYSGVNNNQFTDQTQDVFGVNNITVTSNSVAL 1869
1441 QY GANGSKSVTGEVIGKDDTYFFPAKDGKQVKGQTSVAG--NGRISYYGDSGKRAVS 1492
1870 DB GSN-----SAISAGTHAGTQAKKSDGTAGTTTTAGATGTVKGFAGTAVGAVS 1917
```

RESULT 7

```
US-11-022-562-228
; Sequence 228, Application US/11022562
; Publication No. US20050249742A1
; GENERAL INFORMATION:
; APPLICANT: Ruprecht, Ruth M.
; APPLICANT: Shisong, Jiang
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING
; TITLE OF INVENTION: A CYTOTOXIC T LYMPHOCYTE IMMUNE RESPONSE
```

RESULT 8

```
US-11-013-759-9
; Sequence 9, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN OF MORAXELLA
; FILE REFERENCE: 1038-921MS:1b
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
```

```
; FILE REFERENCE: DFN-043CN
; CURRENT APPLICATION NUMBER: US/11/022,562
; CURRENT FILING DATE: 2004-12-22
; PRIOR APPLICATION NUMBER: PCT/US03/20322
; PRIOR FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: 60/392718
; PRIOR FILING DATE: 2002-06-27
; NUMBER OF SEQ ID NOS: 340
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 228
; LENGTH: 396
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-11-022-562-228
```

```
Query Match 4.1%; Score 322; DB 7; Length 396;
Best Local Similarity 24.9%; Pred. No. 1.1e-10;
Matches 126; Conservative 53; Mismatch 164; Indels 164; Gaps 21;

QY 1033 HSLYVANSKSSGKDYOQAKYGG-----EFLAELKAKYPMFKV-----NMISTGKPIDDS 1081
DB 16 HKVWINLSSSFEYKWNSTEGSDPILVRYLESNKKILQIRIKGILSNKTSFNKMSIDF 75

QY 1082 VKLKQWKAEY-----FNGTNVLRG--VGYVLSDEATGKFTVTKEGNFIPQLTGTGKEK 1133
DB 76 KDIKKLSLGYIMSNFKSFENSENELORDHLGPKIIDNKTYVYDEAS-----K 121

QY 1134 VITGSSDGKGYTGTSGTOAKSAFTVFNNGTNYTFDARGHMTVNSYSPNGKDVYRFLP 1193
DB 122 LVKGLININNSLFYDPDIESNLVTQWTINGKYYFDINTGAASTSYKIIINGKHFY-FNN 180

QY 1194 NGIMLSNAPYIDANGNTLYNSKGQMYKGYTKFDVSETDKDGKESKVVKFPYFTNEGVM 1253
DB 181 NGVMQLGVF-----KG---PDGFYFAPANTQNNIEQAIIVY-----215

QY 1254 AKGVTVIDGFTQYFGEDEGQAKDKLVTFKGTYYFD-----AHTGNGIKDTWRNNGKYY 1309
DB 216 -----QSKELTLNGKYYFDNDSKAVTG-----WQTDGKYY 248

QY 1310 PDAN-GVAATGAQVINGOKLYFNEBDSQVKGKGVVKNADG-----TYSKYNASTG 1368
DB 249 FNLTAEAAATGWTIDGKYYFN-----TNTSIAS- 278

QY 1369 GNVVYAGANGKTVTGAOVINGOHLVFNADGSQVKGKGVVKNADG-----TYSKYNASTG 1422
DB 279 -----TGYTIINGKHFYFNTDGI-MQLGVFKGPNGFYFAPANTDANNIEG 323

QY 1423 E--RLTNEFFTTGDNWYIIGANGKSVTGEVIGKIDDTYFFPAKDGKQVKGQTSVAGN--- 1477
DB 324 QAIRYQNRFLYLHD-NIYFNGNSKAVTGWQTINGNVVYFMPD-----TANAAAGGLPE 376

QY 1478 --RISYYGDSGKRAVSTWIEIQGVY 1502
DB 377 IDGVIYFFGVGVKA-----PGIY 395
```

```
; SOFTWARE: PatentIn Ver. 2.1.1
; SEQ ID NO 9
; LENGTH: 2053
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-9

Query Match
Best Local Similarity 20.3%; Pred. No. 2.3e-09;
Matches 347; Conservative 225; Mismatches 696; Indels 445; Gaps 81;

QY 24 VALATVGLGLSVTSSVSADETDQTKVTQSNSTG---TASLVTSPEATKEADKRTNTKE 79
DB 412 VKLAKELTGL---TSVSA--TNKLTVSTNNNAELQGGGLTFSPITGKTDK--TVVS 463

QY 80 ADVLTPAKETNAVETATTNN--TQATAEAATATTADVAVAAPNKEAVV---TTDAPAVT 135
DB 464 IDGLKFTNDSNSTATGTTTRITKKKIGPAGTNDGVDESKPYLDNEKLVGNSTLNGSLT 523

QY 136 TEKABEQP-----ATVKAEEVN-----TEVKAPPAALKDSEVEAALSILK 174
DB 524 VNNNTGNKQIQVANGIKFATVANNVANTSATVGTARITEEKIGFAGTNDGVDEQAPYLD 583

QY 175 NIKNIDGKYVYVNEDESHKENFAIT--VNGQLLYFGKOCALTSSTSYSTPTGTTNIVDGF 232
DB 584 KERLKVGRVEIITDGSINAGNHKIGLTNG-----TANTDAVTIKQLDKAKP--TLNAGDGI 638

QY 233 SINRAYDSSEASFELIDGLYTDASW---YRPASIIKOGVT---WQASTAEDFRLPLMA 285
DB 639 SINSNNGDLVDSS-----GNITPTVNI SVKTTKLNSTNGTSGNNKFSVSNADHNSLV-- 691

QY 286 WNPVVDVTQVNYLYNYSKVFNLDKAYSSTDKQETLKVAADKDIQK-----IEQKIQAEKSTQ 341
DB 692 -----TAKDLADYLNKV-----NETADSALPSPFKVQNGDNSNNAITVYKOT- 732

QY 342 WLRETIASAVKTPQPNWKENYSKGGEDHLQGGALLVYNDSTRTPWANSDYRLNRAT 401
DB 733 -----NGKTFNLTCLKGEN-----GWNITTN 753

QY 402 NOTGT-----IDKS-----ILBQSDPNHMGDFLL----- 428
DB 754 RATGVTTFGIDQSNGLTTPKLTVGSDTNGNRLVIEQVPSADGNSKTNIIGLSPTLPSTA 813

QY 429 ---ANDVDLSNPVQABQLN--QIHYLMWNG--SIWVGDKDANPDGIRVDADVNDADMLQ 482
DB 814 SPGRNIALGNTIEEKDKSNAASIDDLVNLNAGFNKNGKDKDFVS--TYDTVDFIDGNATT 872

QY 483 LVTNY-----PREYGVNKSBA-----NALAHISV---LEAWSLNDNHYNKDTGCA 525
DB 873 ATVTYDEANQTSKVAYDVNVDEKTIETLTDGNGKKQLGVKTIKLTETSTNGNATTSTDD 932

QY 526 ALAMENKORALLFSLAKPI-KERTPAVSPLYNFTNTTQORDEKT-----WINKDGSKA 579
DB 933 HALVKASDIAGNLNTLAEIHTTKGTANTALQTFVKKYVDENDKADDTNATITVGKDTSG 992

QY 580 -----YNEDGTVKQSTIGKNEYKYGDSAGNVVFIKRAHNNVQDIIAEII 623
DB 993 KVNTLKLGKNGLIDIKTDKDGTV---TFG-INTQSLKAGDSTTL---NNNGLSIKNTAS 1045

QY 624 KKEINPKSGFTITDAEMKQAFEINYKMDLSSDKKYTLNNIPAAVAVMLQNMETTIRVY 683
DB 1046 NFOIQVGADGV-----KFAWNVNGVVGAG-----IDGTRITRDEI 1081

QY 684 GDLTDDGHYMETKSPYDYTI-VNLMSKRIKYSVGGQ-AQRSYWLPTDGMKNSDVLYR 741
DB 1082 G--FTGTNGSLDKSPHLSKDGINAGGKKITNIQSGEIAKNSHDAVTGGKIYDLKTEL-- 1137

QY 742 THEVYTSVRYGKDIM---TANDEGSKY-----SRTSGQVTLVANN----- 779
DB 1138 ENKISHTAKTQNSLHESFVADEQGNFTVSNPYSSYDTSKTSVDITFAGENGITTKVKN 1197

QY 780 --PKMLDQSAKL-----NVEMGKIHAKQVRALLIVGTADGINKFTSDADA1---- 824
```

```
DB 1198 GWRVGIDQTKGLTTPKLTVGNNGKGIVINSQNGQNTIITGLSNTLANVTNDKGSVRTTE 1257
QY 825 AAGYVKETDSN-----GVLTFGAN-----DIKGYETFDMSGGFVAVWVPVQASDNQ 869
DB 1258 QGNIIKDEBKTRAASIVDVLGAGFNLCQNGEAVDFVSTYDITVNFANGNTTAKAVTYDDTS 1317
QY 870 -----DIRVAPSTEAKKEGEL-----TLKATEAYDSQLIVEGFSNFQTIIPDGSPPSVY 917
DB 1318 KTSKVYVDNVDDTTTIEVKDKKLGKVTTLTSTGTGANKFALSNQATGDALVKASDIIVAH 1377
QY 918 TNRKIAENVDLFKSWGVTSEFMAPOFVSADDTGTFDLSVIQNGYAFADRYDLAMSKNNKY- 976
DB 1378 LN---TLSDGIQTAKGASQANNSAGYVDADGNKVI-----YD---STDNKKY 1418
QY 977 -GSKEDLDALKALHAKIGIQAIADWVPDQIYQLPQKEVVTATRTDGAQRKIADAIIDHSL 1035
DB 1419 QAKNDGTVDKTEKVAKDKLVAQAQTPDGLTACQMNKVSINKEQVNDANKK--QGINEDNA 1476
QY 1036 VYANSKSGDYQAKYGGEGFLAELK--AKYPEMPKVMIMSTGKPIDDDSVKLGKQWKAIFYN 1093
DB 1477 FVKGLEKAASDNKTKNAAVTVGDLNVAQTPLTTFAGDTGTAKKLGELTITKGGTD--- 1533
QY 1094 CTNVL-ERGVGVVLSDEATGYFTVTKEGNFIPLQLTGKEKVIITGFSS-DGKGIYFGTS 1151
DB 1534 -TNKLTNNIGVAGTDG---FTVK-----LAKDLTNLNSVNAGGTTKIDEGISFVDAN 1583
QY 1152 GTQAKSAFVTNGNTYYFDPARGHMVTNSEYSPNGKDVYRFLPNGIMLSNAFYDANGNTY 1211
DB 1584 G-QAKA-----NTPVLISA-----NGLDL-----GGKVISNV-----GKGTND 1614
QY 1212 LYNKSGQMYKGYTKFDVSETDKGKESKVKVFRYFTNEG-----VMAKGVTVIDG--- 1262
DB 1615 TDAANVQOLNEVRNLLGLNDNADGNQVNIADIKDPNCGSSSNRTVIKAGTVLGKGN 1674
QY 1263 FTQYFEGEDGFQ-AKDKLVTFKGYTFYDAHTGNGIKDWTW-----DGNANGDLSNVVKTQDKGSKKALLATYNAAGQ 1723
DB 1675 DTEKLATGVGVGVDK-----DGNANGDLSNVVKTQDKGSKKALLATYNAAGQ 1723
QY 1307 WYFEDANGVAATGAQVINGOKLYFNEDGSO---VKG--GVVKNADGTYSKYKEGFGELVT 1361
DB 1724 TNYVTNNPAEAIIDRINEQGIIRFFHVDNGNQBVPVQGRNGIDSSASGKHS-VAIGFQAKAD 1782
QY 1362 NEFF-----TTDGNVWYVYAGANGKTVTGAO-----VINGOHLXFNADGSOVKGGV 1407
DB 1783 GEAAVAIGRQTAQAGNQSIAIGDNAQ-ATGDOSIAIGTGNVWVAGKHSIGAIDPSTVK--- 1837
QY 1408 KNADGTYG--KYNASTGERLTNEF-----FTTGDNNWYIIGANGKSVTGEVKIGDDTYFF 1460
DB 1838 --ADNSYSGVGNNOFTDATQTDVFGVGNNTVTSNVALGSN-----SAISAGTHAGTQ 1890
QY 1461 AKDGKQVKGQTVSAG-NGRISYYGDSKRAVS 1492
DB 1891 AKKSDGTAGTTTTAGATGTVKGFAGQAVGAVS 1923

RESULT 9
US-11-052-554A-217
; Sequence 217, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 217
```

; LENGTH: 726			
; TYPE: PRT			
; ORGANISM: Streptococcus mutans UA159			
US-11-052-554A-217			
Query Match 3.8%; Score 301; DB 7; Length 726;			
Best Local Similarity 31.1%; Pred. No. 3.6e-09;			
Matches 68; Conservative 41; Mismatches 90; Indels 20; Gaps 5;			
QY 1304 NGKWTYFPDANGVAATGAQVINGOKLYFNEEDGSQVKGWVKNADGYSKYKEGFGELVTNE 1363		QY	-GVTWOASTAEDFRPILLMAWPNVDQVNYLYNMSKVFN-----LDAKYSSTDKQETLKV 321
DB 157 NGWYILGADGRNVTSHTIGKTYTFAQDGKQVKGAFADSDGNKHYYDRDSEGWYTR 216		DB	483 IGFSNKAQVVDENKPYL-----DKDKLKVGNSTLNNGLTVANTTIGSNKQ--IQV 531
QY 1364 FFTDGNVYIYAGANGKTYTGAQVINGQHLHYFNADGSQVKGWVKNADGYSKYKNASTGE 1423		QY	322 AAKDIO---IKIBOKIOAEKSTOWLBRETISAFVKTOPQWNKETENTYKSGGGEHLOGGAL 378
DB 217 FVNDQGN-WYLLNNDGVPVTGTSITVNGQSLYFNSDGSQVKGNFVEE-DGSLRYDYKNSGD 274		DB	532 GADGIKFADVNVNVNSNAKFGTTTRITEEIGFADADGKVKDKSPYLDK-----KQLQVGGV 587
QY 1424 RLTFNEFTTCDNNWYIYGANGKSVTEVKIGD-----DTYFFAKDGKQVKGQTVSAGNR 1478		QY	379 LYVND-----SRTPWANSND-----YRRLNRATATNOTGTIDK-SILDEQ----- 415
DB 275 LLKRTSRINGVNYQFDNDGNARAIDKIEVVKTSLVVDSYEFQFSVSKI-----ILEFNHK 330		DB	588 KITKDSGINAGDOKISNVKDATDDTDAVYKQLKQVQDADGALQSFSDINDEKQGEFTIS 647
QY 1479 ISYVYDGSQ-----KRAVSTWIEIQGVVYVYFDKN 1508		QY	416 -----SDPNHMGDFDLLANDVDLSNPVVOAE---QLNQIHYLMWGSIVMGDKD--- 462
DB 331 VTPAVVHAGAMVTTAGVQKILNSYVSNASHVHYFDSS 369		DB	648 NLYSNGNTENTPETITFAGENGISISNDIAKGVKVGIDPINGLITTPKLTGVSDDKDKTKQ 707
RESULT 10		QY	463 -----ANFDGTRV----- 470
US-11-013-759-11		DB	708 LVIEQVASGNDTKNIIRGLSPTLPSITNAGVTRTTEQGNITTSDEDEKSAASIGDILNTG 767
; Sequence 11, Application US/11013759		QY	471 -----DAVDNDADMLQLYTNV-----FREYVGYNKSEA-----NA 501
; Publication No. US20050249747A1		DB	768 FNLKNNNSVGVSTVNTVDFIDGNATTAKVYDETNOTSKVTYDVNVDEKTIETLGDNG 827
; GENERAL INFORMATION:		QY	502 LAHISVLEASLNDNHNDK-----TDGAALAMENKQRLALIFSLAKPIKERTPAVSPL 555
; APPLICANT: Loosmore, Sheena M.		DB	828 KTKNGKVTTLTTTNANGKATNFSTTDNDAL-VNAKDAENLNTLAKEL-HTKGTADT 885
; APPLICANT: Sasaki, Ken		QY	556 YNNTFNWTO---RDEKTOWINKDGSK-----AYNEGTV----- 586
; APPLICANT: Yang, Yan Ping		DB	886 ALQTFKVKDGDATDDETIIVGKDGTONGKTVNTLKLKGENGLTVATNKGDTVFVGINTQS 945
; APPLICANT: Klein, Michel H.		QY	587 -----KQSTIGK-----YNEKYGDASGNVVFIRAHNNVQ---DIAEIIKEINP 629
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE		DB	946 GLKAGDSITLNDKGLSINKNPASNEQIQVGADGVKFAKVDKNGSSTGIDGTSRTKDOI-- 1003
; TITLE OF INVENTION: PROTEIN OF MORAXELLA		QY	630 KSDGETITDAEMKQAFIYNNKDMLS-SDKYTLNNIPAAAYVMLQNMETITRVYVGDLYT 688
; FILE REFERENCE: 1038-921MIS:jb		DB	1004 ---GFTGANGSLDTTKPHLTOKLKVGEVEIINTGINA-----CGKKTINQSGDI-T 1052
; CURRENT APPLICATION NUMBER: US/11/013,759		QY	689 DDGHYNETKSPYYDITIVNLMSRIKYVSGGQORSYMLPTDGMNDNSDELVELYRTNRYVTS 748
; PRIOR FILING DATE: 2004-12-16		DB	1053 QNSNDAVTGGRVYDLKTEL-ESKIN-SAAKTAQNS-----LHEFSVA 1092
; PRIOR FILING DATE: 1999-07-27		QY	749 VRYGKDIWNTANDTEGSKYRTSQVTLVANN-----PKLNLDSAKLVNVEGKIH 798
; NUMBER OF SEQ ID NOS: 32		DB	1093 DEQGNHFTVSNPYSSYDTSKTSQVITFAGENGITTKVKNKGVVRVIGIDQTKGLTTPKLTV- 1151
; SOFTWARE: PatentIn Ver. 2.1		QY	799 ANQKYRALIVGTADG-----IKNFTSD--ADAAAGVYKVTDSNGVLTFFG----- 841
; SEQ ID NO 11		DB	1152 GNNNGKIVIDSKDQNTITGLSNTLANVTNDGAGHALSQGLANDTDKTRAASIGDVLNA 1211
; LENGTH: 2314		QY	842 -----ANDIKGYETFDMSGFVAVMVPVVGASD-----NODIRVAPSTEAK 880
; TYPE: PRT		DB	1212 GFNLQNGEAVDFVSTYDVTDFIDGNATTAKVYDIDTSKTSKVAVDVNVNKNKTIEVTSK 1271
; ORGANISM: Moraxella catarrhalis		QY	881 KEG--ELTLKATEAVDSQLIEGFSNFQTIPOGSDPSVYTRKRIKENVDLFSKWSGVTSPE 938
US-11-013-759-11		DB	1272 KLGVKTTTLTKTSANGNATKFS-ADGDALVKASDIATHLN-TLAGDIOQAK--GASQAS 1327
Query Match 3.8%; Score 299; DB 7; Length 2314;		QY	939 MAPOFVSADGGTFL-DSVIQNGYAFADRYDLAMSNNKYKSKEDLRDLAKLHKGAGIQA 997
Best Local Similarity 19.5%; Pred. No. 2.4e-08;		DB	1328 SSASVYDADGNKVIYDSTDKKYQVNDKQO--VDKN-----KEVAKDLVA-----QA- 1373
Matches 374; Conservative 229; Mismatches 598; Indels 616; Gaps 93;		QY	998 ADMVPD-QIYQLPGKEVVTATRTDGAGRKIADAIIDHSLYVANSKSGKDYQAKYGGEFL 1056
QY 11 KYKKQWHTIIVASVALATVGLSVTSSVSADETQD-----KTVTQSNSGTTAS 60		DB	1374 --QTPDGTLAQMWKVSVINKEQVNDANKK--QGINEDNAFIKGLENAAKDTTKNAATV 1429
DB 239 KAEKG-TIAIGSNAQAINVGAL-----ALGADTRVDLDYGIALGYGSOILNANN----- 287		QY	1057 AELK--AKYPENFKVNMISTGKPIDDSVKLKQWKAIEYFNGTVNL-ERGVGYVLSDEATGK 1113
QY 61 LVTSPEATKADKRTWKADVLTPAKETNAVETATTTTQATAEAAATTATTADVAVAV 120		DB	1430 GDNLVAQTPLTFFAGDTGTTAKKGLGTLTIKGGQTD-----TNKLTDDNNIGVAVAGTDG--- 1482
DB 288 -----NANNKAYV--PEGNGSNTKSKATGNGLFSIGSSTIKRKIINVGA- 330			
QY 121 PNKEAVVTTDPAVTTKEAEQPA-----TVKAEVNVTEVK---APEAALKDSEVEA--- 169			
DB 331 ----GYEDTDVNVVAQLKAIVENLAKKQITFKGDDNGTGVKKLGETLTIKGETQADKLT 386			
QY 170 -----ALSKNKNIDGKYVYVNEEDGSHKENFAITVNGQLLYFGKDGALTS 216			
DB 387 DNNNIGVVDNNTGLKVKLAKNLSGLETVSTKNLTASEK--VTGVS-----GNNTAELQS 439			
QY 217 STYSFTPGTNTI-----VDGFSINNRAYDSSEASFELIDGILTADSWYRPASIIKD- 267			
DB 440 GGLTFTP-TTNASTDKTYVGTDLKFT-----DNSNTALE-----DTTRITKDK 482			

QY 1114 YFTVTKEGNFIPLOLTGKEKVTGPSS--DGKGIYFTGTS--TOAKSAFVTFNGNTYYDA 1171
Db 1483 -FTVK-----LAKDLTNLNSVAGGTRIDEXGISFVDANGQAKANTPVLISANG-----LDL 1532
QY 1172 RGHMTNSEYSPNGKDVRFLPNGLMLSNAFYIDANGNTYLYNSKGMYKGGYTKFDVSE 1231
Db 1533 GGRKISNIGAAVDDNDVNFQFNEVAKT-----VNLNNSQNSGASLFPVVD 1581
QY 1232 TD-----KDGKESKVVKFR-----YFTNEGVMAGVTVIDGTFYFGEDG--FOAKDKLV 1279
Db 1582 ANGKPIGTDGKPKQRAIKGADGKYHAN-----ANGVPV-----DKDGKPIITDADKLA 1629
QY 1280 TFKGXTYYDA-HT-----NGIKDTWNINGKYYFPDANGV----- 1315
Db 1630 NLAAGKPLDAGHQVVASLGGNSDAITLTNIKSTLPQIDTPNTEGNANAGQAQSLPSLSAA 1689
QY 1316 ----AATGAQVIN-GOKL-----YFNEDGSQVKGVVKNADGTYSKYK 1353
Db 1690 QQSNAASVKDLVNLGNFLOTNHNQVDFVKAYDTVNFVNGTGADITS--VRSADGTMNIT 1747
QY 1354 EGFELVITNEFFTGNV-----WYYA-----GANGKTVTGAQVIN-- 1389
Db 1748 VNTALAAATD-----DGNVLKAKDGKFKYKADDLMPNGLKAGKSASDAKTPTGLSLVNP 1803
QY 1390 -----GQHLNFADGSQVKGVVKNADGTYSKYNASTGERLTNEFFTGDNNWYIGAN 1443
Db 1804 AKGSGTGDAVALN---NLKAVFKSKGTTTTTTSVSSDGISIQGK-----DNSSITLSKD 1854
QY 1444 GKSVTGEV-----KIGDDTYFFAKDGKOV----- 1467
Db 1855 GLNVGKVISNVGKTKYDAAVQOLNEVRNLLGLGNAGNDN-----ADGNQVNIADIKK 1910
QY 1468 -----KGQTSAGNRSIYYYGDSGKRAVSTWIRIQPVVYVYFDKNGLA 1511
Db 1911 DPNSGSSNRITKAGTVLGGKGN-----NDTEKLATG-----GVQGVVDKOGNA 1955

RESULT 11
US-11-052-554A-229
; Sequence 229, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 229
; LENGTH: 619
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae R6
US-11-052-554A-229

Query Match 3.5%; Score 280.5; DB 7; Length 619;
Best Local Similarity 27.8%; Pred. No. 3.8e-08;
Matches 81; Conservative 20; Mismatches 99; Indels 91; Gaps 10;
QY 1207 NGNTYLYNSKGOMYKGGYTKFDVSETDKDGKESKVVKFRYFTNEGVMAGVTVIDGFTQY 1266
Db 408 NCMWTFYNTDGMATGW-----LQNG-----SWYLYNSGAMA-----TGWLOY 447
QY 1267 FGEDGFQAKDKLVTFKGTYYTFAHTGNGIKDT-WRNINKRWYTFDANGVAATGAQVING 1325
Db 448 -----NGSWYLYNA---NGAMATGNKAVNGSWYLYNANGAMATGWLQYNG 489
QY 1326 QKLYFNEDGSQVKGVVKNADGTYSKYKEGFGELVTNEFFTGDGNVWYVYAGANGKTVTCA 1385

Db 490 SWYLYNANGAMATGWAKVNGS-----WYLYNANGAMATGW 524
QY 1386 QVINGQHLNFADGSQVKGVVKNADGTYSKYNASTGERLTNEFFTGDNNWYTYGANGK 1445
Db 525 LQYNGSWYLYNANGAMATG-----WAKVNGS-----WYLYNANGA 559
QY 1446 SVTGEVKIGDDTYFFAKDGKQVKGTVSAGNRSIYYYGDSGKRAVSTWIE 1496
Db 560 MATGWVKDGTWYLYLEASGAMKASQWPKVSDK--WYVYNGLGALAVNTTVD 608
RESULT 12
US-11-052-554A-231
; Sequence 231, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 231
; LENGTH: 701
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae R6
US-11-052-554A-231

Query Match 3.5%; Score 276.5; DB 7; Length 701;
Best Local Similarity 29.3%; Pred. No. 7.4e-08;
Matches 78; Conservative 22; Mismatches 109; Indels 57; Gaps 7;
QY 1232 TDKGKESKVVKFRYFTNEGVMAGVTVIDGFTQYFGEDGFOAKDKLVTFKGTYYFDDAH 1291
Db 480 TPKTGMKQENGWYFYNTDGMATGWLNQNGSWYLYNANGAMATGWLQNG-NGSWYLYNA- 537
QY 1292 TNGIKDT-WRNINKRWYTFDANGVAATGAQVINGQKLYFNEDGSQVKGVVKNADGTYS 1350
Db 538 --NGSMATGWLQNGSWYLYNANGAMATGWLQYNGSWYLYNSGAMATGWLQYNGS---- 591
QY 1351 KYKEGFGELVTNEFFTGDGNVWYVYAGANGKTVTGAQVINGQHLNFADGSQVKGVVKNA 1410
Db 592 -----WYLYNANGDMATGWLQNGSWYLYNANGDMATG----- 624
QY 1411 DGTYSKYNASTGERLTNEFFTGDNNWYTYGANGKSVTGEVKIGDDTYFFAKDGKQVKGQ 1470
Db 625 ---WLQYNGS-----WYLYNANGDMATGWVKDGTWYLYLEASGAMKASQ 665
QY 1471 TVSAGNRSIYYYGDSGKRAVSTWIE 1496
Db 666 WPKVSDK--WYVYNGSGALAVNTTVD 689
RESULT 13
US-10-873-528-184
; Sequence 184, Application US/10873528
; Publication No. US20050276814A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129W
; CURRENT APPLICATION NUMBER: US/10/873,528
; CURRENT FILING DATE: 2004-06-23
; PRIOR APPLICATION NUMBER: US/09/769,787

; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 184
; LENGTH: 744
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-873-528-184

Query Match 3.4%; Score 267; DB 6; Length 744;
Best Local Similarity 27.1%; Pred. No. 2.7e-07;
Matches 73; Conservative 28; Mismatches 106; Indels 62; Gaps 6;

QY 1232 TDGDKESKVVKFRYPFTNEGVMAKGVTVIDGFTQYFCEGDFQA-----KDKLVTFKGTYY 1287
DB 523 TPTKWKQENGMYFYNTDGSMAIGMLQNNQNGWYLLNANGMATGWYKD-----GDTWY 576
QY 1288 FDAHTGNIGKIDTWRNNGKYYFDANGVAATGAQVINGQKLYFENEDGSQVKGWVKNADG 1347
DB 577 YLEASGAMKASQFKVSDKYYVNSNGAMATGMLQYNGSWYLLNANGMATGHLQYNGS- 635
QY 1348 TYSKYKEGFGELVTNEFFTTDGNWYYAGANGKTVTGAQVINGOHLFYFNADGSQVKGWV 1407
DB 636 -----WYLLNANGMATGWAKVNGSWYLLNANGAMATG--- 668
QY 1408 KNADGTYSKYNASTGERLTNEFTTGDNNWYIYGANGKSVTGVKIGDDTYFFAKDGKQV 1467
DB 669 -----WAKVNGS-----WYLLNANGMATGWVXOGDTWYILEASGAMK 706
QY 1468 KGTVSAGNGRISYYGDSKRAVSTWIE 1496
DB 707 ASQWFKVSDK--WYYNGLGALAVNTVD 733

RESULT 14
US-11-052-554A-260 | Application US/11052554A
; Sequence 260, Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 260
; LENGTH: 1647
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes MGAS8232
US-11-052-554A-260

Query Match 3.2%; Score 255.5; DB 7; Length 1647;
Best Local Similarity 18.2%; Pred. No. 3.5e-06;
Matches 322; Conservative 222; Mismatches 607; Indels 621; Gaps 81;

QY 1 MENKIHKLHKVKKQWTVAVASVALATVGLGLSVTSSVADETDKTVTQSNQSTTAS 60
DB 1 MEKQRFSLRKYSKGTGFSVLIGSVFL-----MMTTVAADE-----LS 38
QY 61 LVTSPEAT-----KEADKRTNKREADVLTAKETNAVETATT---TNTQATAAATATTATTA 113
DB 39 TWSEPTITNHTQOQOHLNTELSSAESKQDTSQITPKTNREKEQPQGLVSEPTTELA 98

QY 114 DVAAVAVPNKEAVVTTDAPAVTTEKAEBOAPATVKAEVVNTFV-----KA 157
DB DTDAAPMAN-----TGPDATQKSASLP-----VNTDHDWVKTKGAWDKYKQG 144
QY 158 PEAAKDXSEVEAALSILKNIDGKYYVNEGSHKE-----NFAITVNGQLLYFGKDGA 212
DB 145 KVVAVIDTGDPAHQSMRISDVSTAKVKGBEDMLAQKAAGINYGSWINDKVVVF----- 198
QY 213 LTSSTYSFTPGTTNIVDG-FSINNRAYDSSEASFELIDGILYLTADSWYRPAS-----I 264
DB 199 -----AHNVYNSDNIKENQFDFEDWENFEFAEAPKAIKKHKIYRPQSTQAPKETV 253
QY 265 IKDGVTTQASTABEDPRPLMAMWPNVDQVNYLYNYSKVFNLDAKYSTSDTKQBTLKVAAK 324
DB 254 IK-----TEETDGSHDID-----WTQDDDTKYESHGMHVTGIVAGNS-----KEAATATGER 300
QY 325 DIQIKLEQIQAEK--STOWLRETIISAFVKTQP-----QWNKETENYKGGGEDH 372
DB 301 FLGIAPEAQVPMFMRVFPANDVMSGAESLFKAIEDAVAGADVINLSLGTANGAQLSGSKP 360
QY 373 LQ-----GGALLYVNDSTPWA-NSDYRRRLNRATNOTGTIDKSI--- 411
DB 361 LMEATEKAKKAGVSVVVAAGNERVYGSDDHDDPLAINDPDLGLVGSPTGRTPTSVAAINSK 420
QY 412 -----LDEQSDPNHMGGFDFLLANDVLSN--PVVQAEQLNQIHYLMNWSIVM 458
DB 421 WVIQRLMTVKELENRADLNHGA---IYSESVDKNIKDSLGLYDKSHQPAY-----VK 470
QY 459 GKDANFDGIRV-DAVDNVDADMLQY----- 484
DB 471 ESTDAGYKAQDVKDKIALIERDPNKTYDEMIALAKKHGALGVLI FNNKPGQSNRSMRLTA 530
QY 485 -----TNYPREYVG-----VNKS---EANALAHISLVEAWSL--- 513
DB 531 NGMGIPSAFISHEFGKAMSQNLNGGTGSLFSPSVSKAPSKQGNENHFS---NWGLTSD 587
QY 514 -----NDNHNKTDGAALAMENKORLALLFSLAKPIKERTPAVSP- 554
DB 588 GYLKPDITAPGGDIYSTYNDNHYGSOT-GTSMA---SPQIAGASLLVQVLEKTPQNLPK 643
QY 555 -----LYNNTFTTQDEKTDWINKDGSKAYNEDGTGVKQSTIGKY---NEKYGD 600
DB 644 EKIADIVKMLMSNAQIHVNPETKTTSPROQGAGLLNIDGAV---TSLYLYVTGKDNYS 700
QY 601 AS-GNVVFTRAHDNNVQDII-----AEIIKKEINPKSDGFTTTDAEMK--QAFEI- 647
DB 701 ISLGNITDWTFTDVTHNLSNKDKTLRYDTELLTHDVPKGRFTLTSLSKTYQGEVT 760
QY 648 -----YNKQML-----SSDKKYTLNNIP-AAYAVML 672
DB 761 VPANGKVTVRVTMDVVSQFTKELTKQMSNGYYLEGFVRPRDSQDDQLNRVNI PVPVFGKQF 820
QY 673 QNMEITR-VYYGDLTYDDGHYMETKSPPYDTIVNLMKSRIKYVSGGQOQRSWLPTDQK 731
DB 821 ENLAVAEESIYRLKSQKGTGFYFDESQPKDDIYVG-----KHFTG-----LVTLGS 866
QY 732 MDNSDVELYRTNEVTSVRY-----GKIDMTANDTGSKYSRSTSGQVTLVANNPKLNQDS 787
DB 867 ETNVSTKISDNGLHTLGTFFKADGRFILEKN-AQGN-----PVLAISPDGNNQDFA 918
QY 788 AKLANVEMGKIHANQYRALIVGTAD-----GIKNFTSDADATAAGYVKET 832
DB 919 AFKGVFLRYK---QGLKASVYHASDKEHKHKNPLWVSPESPKGDKNFNSDIFRAKSTTLTGT 975
QY 833 DNGVLTFGANDIKYETFDMSGGFVAVWVVPVGASQNDIRVAPSTEAKEGELT-----LK 888
DB 976 AFSGSLTGAELPDGYHYVVSYY-----PDVVGAQRQEMTFDMILD 1017
QY 889 ATEAEDSOLLYBGFNS-FQTIIP-----DGSDFSVVYTNRIAE--NVVDLFSKSGVTSF 937
DB 1018 ROKPVLVSQATFPFETNRRFKPEPLKDRGLAGVRKDSVFLERKDNKPYTIVINDSYKYSV 1077
QY 938 ENAPQFVSAD-DGTFDLSVIQNGYAFADRYDLAMSK---NNKYGSKEDLRDALKALHKAG 993

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 19:15:50 ; Search time 31.3322 Seconds
(without alignments)
4661.567 Million cell updates/sec

Title: US-10-797-821-40
Perfect score: 7928
Sequence: 1 MENKIHYKLHKVKQWVTIA.....PGVYVFDKGLAYPRVLN 1518
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7928	100.0	1518	2 A44811	glucosyltransferas
2	3743	47.2	1599	2 S22737	glucosyltransferas
3	3515	44.3	1431	2 A45866	dextranucrase (EC
4	3296.5	41.6	1592	2 A38175	glucosyltransferas
5	3196.5	40.3	1475	2 B33135	gtfB protein precu
6	3155	39.8	1577	2 T30858	glucosyltransferas
7	3070	38.7	1375	2 JT0345	dextranucrase (EC
8	3067	38.7	1449	2 T30857	glucosyltransferas
9	3056	38.5	1449	2 T30552	glucosyltransferas
10	2804	35.4	1365	2 A41483	glucosyltransferas
11	2722	34.3	1290	2 JCS473	dextranucrase (EC
12	2638.5	33.3	1508	2 T31098	probable dextranu
13	582.5	7.3	2817	2 B97033	uncharacterized pr
14	522	6.6	2710	2 A37052	toxin A - Clostrid
15	419	5.3	2364	2 I40884	cytotoxin L - Clo
16	405	5.1	648	2 S10869	enterotoxin A - Cl
17	394	5.0	2178	2 S55805	alpha-toxin - Clo
18	367	4.6	2367	2 S70172	toxin B - Clostrid
19	366	4.6	2366	2 S10317	glucan-binding pro
20	344.5	4.3	563	2 A37184	probable peptidogl
21	319.5	4.0	1806	2 AF1717	cell wall-associat
22	300.5	3.8	2334	2 S32920	AAS surface protei
23	297	3.7	1463	2 T30290	hypothetical prote
24	294.5	3.7	1959	2 AG1085	cytotoxin RTX homo
25	285	3.6	1829	2 S35027	surface protein ps
26	280.5	3.5	619	2 A97887	surface protein ps
27	280.5	3.5	619	2 A41971	hypothetical prote
28	278.5	3.5	1946	2 AE1449	cell wall-associat
29	278.5	3.5	2167	2 AF1489	

ALIGNMENTS

RESULT 1

A44811

glucosyltransferase (EC 2.4.1.-) I - Streptococcus salivarius

C:Species: Streptococcus salivarius

C>Date: 31-Dec-1993 #sequence revision 31-Dec-1993 #text_change 09-Jul-2004

C:Accession: A44811; S22726; S28809

R:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.

J. Gen. Microbiol. 137, 2577-2593, 1991

A;Title: Molecular characterization of a cluster of at least two glucosyltransferase ge

A;Reference number: A44811; MUID:92148377; PMID:1838391

A;Accession: A44811

A;Molecule type: DNA

A;Residues: 1-1518 <GIP>

A;Cross-references: UNIPROT:Q00600; UNIPARC:UPI000000BEF31; EMBL:Z11873; NID:947526; PID:

A;Note: sequence extracted from NCBI backbone (NCBIN:81050, NCBIP:81052)

C;Genetics:

A;Gene: gtfJ

C;Keywords: glucosyltransferase; hexosyltransferase

F:1307-1326/Domain: cpl repeat homology <CP4>

Query Match	100.0%	Score	7928;	DB	2;	Length	1518;
Best Local Similarity	100.0%	Pred. No.	0;				
Matches	1518;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	MENKIHYKLHKVKQWVTIAVASVALATVILGSLVTTSSVSADETQDKVTQSNSTTAS	60				
Db	1	MENKIHYKLHKVKQWVTIAVASVALATVILGSLVTTSSVSADETQDKVTQSNSTTAS	60				
Qy	61	LVTSPKATKEADKRTNTEADVLTTPAKETNAVETATTTNTQATAEAATTTATTADVAVA	120				
Db	61	LVTSPKATKEADKRTNTEADVLTTPAKETNAVETATTTNTQATAEAATTTATTADVAVA	120				
Qy	121	PNKEAVTTDAPAVTTEKAEQATVKAFFVNVTEVKAPKALDKSEVAALSILKNIKNID	180				
Db	121	PNKEAVTTDAPAVTTEKAEQATVKAFFVNVTEVKAPKALDKSEVAALSILKNIKNID	180				
Qy	181	GKYYVNEGSGHKNFAITVNGQLLYFGKDGALTSSSTYSFPTGTTNIVDGFSINN	240				
Db	181	GKYYVNEGSGHKNFAITVNGQLLYFGKDGALTSSSTYSFPTGTTNIVDGFSINN	240				
Qy	241	SSEASFELIDGVLTDADSVTRPASIIKDGVTWQASTAEADPRLLMAWPNVDQVNYLNYM	300				
Db	241	SSEASFELIDGVLTDADSVTRPASIIKDGVTWQASTAEADPRLLMAWPNVDQVNYLNYM	300				
Qy	301	SKVFNLDAKYSSTDKQETLKVAADQIKIEQIKIAEKSTQWLRETISAFVTKQPNWKE	360				
Db	301	SKVFNLDAKYSSTDKQETLKVAADQIKIEQIKIAEKSTQWLRETISAFVTKQPNWKE	360				
Qy	361	TENYSGGGEDHLOGGALLYVNDSTPWNYSYRLNRTATNTQGTIDKSIDDEQSDPNH	420				
Db	361	TENYSGGGEDHLOGGALLYVNDSTPWNYSYRLNRTATNTQGTIDKSIDDEQSDPNH	420				

probable peptidogl
choline binding pr
hypothetical prote
muramidase-release
fibrinogen-binding
hypothetical prote
iron-regulated pro
pneumococcal surfa
150K mating aggreg
aggregation protei
probable peptidogl
hypothetical prote
hypothetical prote
filamentous hemagg
probable peptidogl

Db 511 EHNTAALSDNGRLRLSLVHGLTRPVTKNGTCARNASMKDLINGGYFGLSNRAEVTSY--- 567
Qy 575 DGSKAYNEDGTVKQSTIGKNEYKYGDSAGNYVFIIRAHNNVQDIIAIEIIKKEINPKSDGF 634
Db 568 -----DQLGPAT--YLFVRAHDSEVQTVIADIISKIDPTDGF 604
Qy 635 TIIDAEMKQAFIYNKMDLSDDKYITLNNIIPAAVAVMLQWETITRVYYGDLTYDDGHY 694
Db 605 TFLDQLKQAFDYNADMLKVDKEYTHSNIIPAAVALMLQTMGAATRVYYGDLTYDNGQYM 664
Qy 695 ETSPYVDTTIVNLMSKIKVYSGQAQRSLPTDGRWMSDVVELYRTNEVTSVSRGKD 754
Db 665 AKSPFDDQITLLKAPKPVAGQTSYIHLNLAGDVSSAKD-----NKEVLVSVRGQD 719
Qy 755 IMTANDTEGSKYRTSQVTLVANNPKNLIDQSAKLNVEMGKIHANQYRALIVGTADGI 814
Db 720 LMSKTDTEGGKYGRNSGMLTLIANNPDLKLADGETITVNMGAHQNAQYRPLLSTEGKI 779
Qy 815 KNFTSDADATAAGYVKETDSNGVLTFGANDIKGYETPDMGFGFVAVMVPVPGASNDQDIRVA 874
Db 780 VSSLNDS--TKIVKYTDAQGNLVTADBEIKGFKTVDMSGYLSVWVPVPGATDDQNVLAK 836
Qy 875 PSTEAKEGELTLKATEAYDSQLIYEGFSNFOTIPDGSDPVSVTNRKIAENVDLFKSWG 934
Db 837 PSTKAYKEGDKVYSSAALAEQVIEGFSNFQFV--KEDSQYTNKLIANADLFKSWG 894
Qy 935 TSFEMAPQFVSADGTFPLDSVIONGYAFADRYDLAMSNNKYGSKEDRLDALKALKHAGI 994
Db 895 TSFEIAPQYVSSKDGTFPLDSIIENGYAFTDRYDFAMSKNNKYGSKEDRLDALKALKHAGI 954
Qy 995 QAITADVDPDIYQIPGKEVVTATRTDGAGRKIADAIIDHSLYVANSKSGKYOAKYGG 1054
Db 955 QVIADVVPDQLYTLPGKEVVTATRTDTHGKVLDDTSLVNKLYVTNTKSSGNDPQAYGGA 1014
Qy 1055 FLAELAKAPPEMFKVMNISTGKPIDDSVKLKQWKAEEVNGTNYLERGVGVVLSDEATK- 1113
Db 1015 FLDKQLKLPFELFEVMEASGKIDPSVKIKQWEAKYFNGTNIQKRGSDVVLSD---GKL 1071
Qy 1114 YFTVTKEGNFIPLQLTGKEKVIITGSSDGKGIYFTGTSQAKSAFVTFNGNTYYFDARG 1173
Db 1072 YFTVNDKGFPLPAALTDGKAKTGFAYDGTGVYIYTTSGTQAKSQFYVNGKQYFNDKG 1131
Qy 1174 HMVTNSEYSPNGKDVTRFPLNGIMLSNAFYIDANGNTYLYNSGQYKGYTKFVDS 1233
Db 1132 YLVT-GEQITDGSN-YFFLPNGWMTDGVVRKNAKQSLVYKSGKLTQTGTGWK-EVTVKD 1188
Qy 1234 KDGKESVKVFRYFTNEGVMKAVTVIDGTQYFEGEDGFOAKDKLTFK----- 1282
Db 1189 DSGKEBK--YQYFFKGGIMATGLTEVGEKERYFYDNGYQAKGVFPTKDGHLMPFCGDS 1246
Qy 1283 ----- 1282
Db 1247 GERKYSPPEDGNWYYANDKGYVATGFTKVKGQNLVFNKGVQVKNRPFQVGDAITYAN 1306
Qy 1283 -----GKTYFFDAHTNGIKDTRWNINGK- 1306
Db 1307 NEGDLVLRGAQTINGDELFPDESQKQVGBFVNNPDGTTYSYDAITGVKLVDTSLVVDGQT 1366
Qy 1307 -----WYFPDANGVAATGAQVINGCKLYFNEGDSQVKGVV 1342
Db 1367 FNVDAKGVVTKAHTPGFYTTGDNWNFYADSYGRNVTGAQVINGQHLVFDANGRQVKG 1426
Qy 1343 KNADGTSYKKEGFGSLVTNEPFTTDGNVWYYAGANGKVTGAQVINGQHLVFNADGSOV 1402
Db 1427 TWTDSRSFYHWTGDKLVSTPATGCHDRYVADRGVNTVGAQVINGQHLVFDGDKQV 1486
Qy 1403 KGVVKNADGTSYKSNASTGERLTNEFTTGDNNWYIYGANGSKVTEGVKIGDDTYFFAK 1462
Db 1487 KGAFATNANGRSYHWTGNKLVSTFTSGDNNWYIYADAKGEVWVVEQTINGQHLVFDQ 1546
Qy 1463 DGOVKQGVTSAGNRISYYGDSGKRAVSTWIEIQPVYVFPDKNGLAY 1512
Db 1547 TGQVKVGATATNPDGSISSYYVHTGEKAINRWVKIPSGQWYFNAQKG 1596

RESULT 3

A45866

dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans

C;Species: Streptococcus mutans

C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Mar-2004

C;Accession: A45866

R;Honda, O.; Kato, C.; Kuramitsu, H. K.

J. Gen. Microbiol. 136, 2059-2105, 1990

A;Title: Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the glucosyl

A;Reference number: A45866; MUID:91100958; PMID:2148600

A;Accession: A45866

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1431 <HON>

A;Cross-references: UNIPARC:UPI000017AC5C; GB:M29296

C;Keywords: glycosyltransferase; hexosyltransferase

F;181-201/Domain: cpl repeat homology <CP1>

F;1127-1146/Domain: cpl repeat homology <CP2>

F;1192-1211/Domain: cpl repeat homology <CP3>

F;1257-1276/Domain: cpl repeat homology <CP4>

F;1277-1297/Domain: cpl repeat homology <CP5>

F;1321-1340/Domain: cpl repeat homology <CP8>

F;1341-1361/Domain: cpl repeat homology <CP6>

F;1385-1404/Domain: cpl repeat homology <CP7>

Query Match 44.3%; Score 3515; DB 2; Length 1431;

Best Local Similarity 48.1%; Pred. No. 3.4e-156;

Matches 724; Conservative 246; Mismatches 436; Indels 100; Gaps 31;

Qy 1 MENKIHKLHKVKQWVTIIVASVALATVLGGLSVTSSVSAD---ETQKTVTQ-SNSG 56

Db 1 MEIKRYKHKVKHVTAVAS--GLITL--GTTILGSSVSAETEQTSDKVVTKSEDD 57

Qy 57 TTASLVTSPKATKDKRNTKEADVLTPAKETNAVETATTNTQTAEAAAT-ATTADV 115

Db 58 KAASESSQTDAPRTKQAQTEQTAQ-----SQANVADTSTSIKETPSQNTTQANSDDK 112

Qy 116 AVAAVNPKEAVVTDAPAVTTEKAEOPATVRAEVVNTVEKA-----PEAALKDSEVEA 169

Db 113 TVTNTKSEEAQTSEE-----RTKQSEEAQTASSQAL-TQAKAELTKQRTAAQENKPNVD 167

Qy 170 ALSLNKINIDGKYVYVNEGDSHKENFALTAVGQLLYFGKD-GALTSSTSYSTPGTTNI 228

Db 168 LAAPVNVKQIDGKYIYIGSDGQPKKNFALTAVNKVLYFDKNTGALTDTSQYQFKQSLTKL 227

Qy 229 VDGFSLNNRAYDSSSEASFELIDGYLTADSWYRPASIIKDGVTWQAETAEDFRPLLMWWP 288

Db 228 NNDYTPHNQIVNFPENTSLTDINYYVTADSWYRPKDIKNGKTWTASSESDLRPLLMWWP 287

Qy 289 NVDTQVNYLVNYSKV-FNLDAKYSTDKQETLKVAAKDQIKIEQKIQAEKSTQWLRETI 347

Db 288 DKQTQIAYLVNYYNQGLGTGENYTADSSQESLNLAAGTVQVKIETKISQTOQTQWLDRDI 347

Qy 348 SAFVKTPQOWNKETENYKSGGGEHLQGGALLVYVNDSTRTPWANSDYRRLNRATNTGTI 407

Db 348 NSFVKTPQWNNSQTESDTSAGEKDLQGGALLYSNSDKTAYANSYRLNRPTFTSQTGK- 406

Qy 408 DKSILDEQSDPNHMGGFDFLLANDVLSNPVVQAEQLNQIHYLMNMGSIVMGDKDANFDG 467

Db 407 -----PKYFEDNSSGGYDFLLANDIDNSNPVVQAEQLNHLVLMNYSIVANDPEANFDG 461

Qy 468 IRVDAVDNVDAADMLQLYTNYFREYGVGNKSEANALAHISVLEAWSLNDHNDKDTGAAL 527

Db 462 VRVDAVDNVNADLLQIADSLKAHYGVDKSEKNAHLHLSLEAWSNDPQYNKDKTKGAQL 521

Qy 528 AMENKORALLFLSAKPIK-----ERTPAVSPLYNNTFNTTQORDEKTDWINKGSKAY 580

Db 522 PIDNKLRLSLLYALTRPLEKDNKNEIRSGLEPVTISLN----- 562

Qy 581 NBDGTGVKSTIGKNEYKYGDSAGNTVFIIRAHNNVQDIIAIEIIKKEINPKSDGFTTDAE 640

Db	753	HSDQE--AAGLVRYTNDRGELIFTAADIKGIANPQVSGYLGVWVPVGA--LIIKFAIR	807
QY	878	EAKKEGEL-TLKATEAYDSOLIEGSPNFQTTIPDSDPSPVYNNRIKAENVDLIFKSWGVT	936
Db	808	LARPHQOMASVHQNAALDSRVMEGFSNFQAF--ATKBEYTNVVIKAVNDKFAEWGVT	865
QY	937	FEMAPQVSADGTFLDSVIQNGYAFADRYDILAMSKNNKYGSKEDLRDALKALHKAGIOA	996
Db	866	FEMAPQVSSDGSFLDSVIQNGYAFDRYDILGSKPNKYGTADDLVKAIALHSGIKV	925
QY	997	IADWVPDQIYQLPGKEVWTATRTDGAGRIADAIIDHSLYVANSKSSGKDYOAKYGEFL	1056
Db	926	MADWVPDQYAPPEKEVWTATRVDTKPTVAGSQIKNTLYVVDGSSGKQQAQYGGAF	985
QY	1057	AEIKAKYPEMKVNMISTGKPIDDSVKLQKWAEBYFNGTNVLERGVGYVLSBATGYFT	1116
Db	986	EELQAKYPELFARKQISTGVPMDDPSVKIKQWSAKYFNGTNILRGAGYVLKQDQATNYFN	1045
QY	1117	VT--KEGNEIPLQLTGKEKVIITCFSSDGKGIYFGTSGTOAKSAFVTFNGNTYYPDARGH	1174
Db	1046	ISDNKEINFLPKTLNLLQDSQV-GFSYDGGKYVYISTSGYQAKNTFISEGDKWYFDDNGY	1104
QY	1175	MYTNSYSNPGKDVYFLPNGLMISNAFYIDANGNTLYNSKGOMYKGGYTKFPDVSETDK	1234
Db	1105	MTGAQ-SINGVNY-FLSNGIQLRDAILKNEDETYAYTNGDGRYENGYYQF-----	1155
QY	1235	DGKESKVKRYPTNEGVMAGVTVIDGFTQYFGEDGFOAKDLV-TFGKTYTFYDAHTG	1293
Db	1156	--MSGV--WRHF-NGGMSVGLTVIDGQVQVFDGMYQAKGFVTTADGKIRYFDKQSG	1209
QY	1294	NGIKDTW-RNINKWYFPDANGVAATGAQVINGOKLYFNEGDSOVKGGVVKNAADTYSKY	1352
Db	1210	NMYNRFIENEGKWLVLGEGDAAVTGSQTINGHLYFRANGVQVKGEFVDHHGRISYY	1269
QY	1353	KEGFGHLVNEFTTDCGNWYTAGANGKTVTAQVINGQHLVFNADGSOVKGGVKNADG	1412
Db	1270	DGNSGDOIENRFRVNAQOGWFFDNGNYAVTGARTINGQLLYFRANGVQVKGEFVTDRYG	1329
QY	1413	TVSKYNASTGERLTNEBFTTGDNNWYIIGANGKSVTGVRKIGDVTYFFPAKQKQVQTV	1472
Db	1330	RISYDGNSGDOIENRFRVNAQOGWFFDNGNYAVTGARTINGQHLVFRANGVQVKGEFV	1389
QY	1473	SAGNGRISYYGDSGKRAVSTWIEIQPVYVVFKNGLA	1511
Db	1390	TDHGRISYYDGNSGDOIENRFRVNAQOGWFFDNGNYA	1428
RESULT 6			
T30858			
Glucosyltransferase - Streptococcus salivarius			
C:Species: Streptococcus salivarius			
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004			
C:Accession: T30858			
R:Simpon, C.L.; Giffard, P.M.; Jacques, N.A.			
Infect. Immun. 63, 609-621, 1995			
A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for pri			
A:Reference number: Z20909; MUID:95122197; PMID:7822030			
A:Accession: T30858			
A>Status: preliminary;			
A:Molecule type: DNA			
A:Residues: 1-1577 <SIM>			
A:Cross-references: UNIPROT:Q55265; UNIPARC:UPI0000088087; EMBL:L35928; NID:g662380; PID			
C:Genetics:			
A:Gene: gtfm			
Query Match			
Best Local Similarity 39.8%; Score 3155; DB 2; Length 1577;			
Matches 687; Conservative 243; Mismatches 494; Indels 194; Gaps 33;			
QY	1	MENKIHKLHKVQKQWTVIIVASVALATVLGGLSVTTSSVSADETODKT---VTQNSGT	57
Db	1	MENKVRFLKHKVQKQWTVIIVTSLMVALAGSLLAQQKVEADETSPAPNGDGLQLQLEDG	60

QY	58	TASLVTSPEATKEADKRTNTKEADVLTPAKETNAVETATTTTNTQATAEAAT-----	108
Db	61	TASLVTTTITVTEQASQAQSVSAVATASVSHETSFQAATSAVSOEATAQAQTSPVASQEVA	120
QY	109	-----TATTADVAAVAPNKEA-----	139
Db	121	VSSQTQSSQOEFTTTEQVSOQOTSQVAGQTSQAQSTPSPVTEQARPRVLTNAAPAIATRAA	180
QY	140	BEQ-----PATVKAEV-----	159
Db	181	DSIRINARNRNTNITITAGTTPNVTTIITGPNTPKPNVTVTSPNGTRPNVTIIVTQNPQN	240
QY	160	AALKDSEV-----EAALSLEK-----	205
Db	241	KPVQSPQSPQPNKPVQPNQPSLDYKPVASNLTKIDGKYIV-ENGWVKNAIAELDGRILY	299
QY	206	YFGKDGALTSSTSYSTFTPGTTNIVDGFSSINNRAYDSSEASFEILDGILTADSWYRPASTI	265
Db	300	YFDEGAMVDQSKPLYRADAIIPNNSIYAVYQAYDTSSKSFHLDNFLTADSWYRPQIL	359
QY	266	KDGVQWQASTAEDFRPLMAMWPNVDTVNLYNYSKVFNLDKAYSSTDKQETLKVAAKD	325
Db	360	KDGKNTASTEKDYRPLMTWMPDKVTQVNYLNYSQQGFGNKTYTTDMMSYDLAAAAET	419
QY	326	IQIKLEQKIOAEKSTOMLRETISAFVKTPQPKNKETENYSKGGEDHLOGGALLYVNDSR	385
Db	420	VORGEBEIRIGRENTTWRQLMSDFIKTPQGNWSESD-NLLVGKDHLOGGALTFLNNSA	478
QY	386	TPWANSDYRRLNRTATNOTGT----IDKSILDEQSDPNHMGGFDFLLANDVDLSNPVQA	441
Db	479	TSHANSDFELMNRFTNTQGTGRKYHIDRS-----NGGYELLANDIDNSNPAVQA	528
QY	442	EQLNQIHYLMNWSIVMGDKDANFDGIRVDAVDNVDADMQLQTYNYFRYYGVNKSSEANA	501
Db	529	EQLNWLHYTNIGSILGNDPSANFDGVRIDAVDNVDADILLQASDYFKEKYRVADNEANA	588
QY	502	LAHISVLEAWSLNDHNHNDKTDGAALAMENKORLALLSLAKPIKERTPAVSPLYNNTFN	561
Db	589	IAHLSILEAWSYNDHOYKNTKGAQLSDINPLURETLTTLTFLRKSNR-GLSERVITNSLN	647
QY	562	TTQRDEKTDWINKDGSKAYNEDGTVKQSTIGKYNBKYGDASGNVYFIRAHDNNVDIAE	621
Db	648	NRSSE-----QKHTPRDANYIFVRAHDSFVQAVLAN	678
QY	622	IIKKINPKSGDFTTIDAEKMQAPEIYNKDMLSKKYTYTLNIPAAIYAVMLQMETITRV	681
Db	679	IIKQINPKTDGFTTMDLQKQAFELYNADIAKADKKYQVNIIPAAIYATMLTKDSITRV	738
QY	682	YVGDLVTDGHHVWETKSPYDITVNLKMSRIKYVSGQAQRSYWLPTDCKMDNSDVVELY	741
Db	739	YVGDLFTDDGQYMAEKSPYINADILLARARIKYVAGGQ-----DMKVTK	782
QY	742	TN--EVTYSRVYKGDIMTANDTEGSKYSRTSQVTLVANNPKLNLDQSAKLAVNEMKIIHA	799
Db	783	LNGYIEMSSVRVKGAEAEANQL-GTAETENQGMVLVTANRPMKLGANDRLVNMGAHK	841
QY	800	NQKYRALLIVGTADGINKFTSDADATAAGVYKKTSDNGVLTFPGANDIKGVETEDMSGFVAV	859
Db	842	NQAYRPLLLSKSTGLATYLLKQSD-VPAGLVRVYTDNQGNLTFTADDIAGHSTVEVSYLAV	900
QY	860	WYPCVGSASNODIRVAPSTAEKKEGELTLKATEAYDSOLIEGFSNFQTIIPDGSDDPSVYTN	919
Db	901	WYPCVASENQDARTKAS--STKKGQVFPSSAALDSQVIYEGFSNFQDFV--KTPSQYTN	956
QY	920	RKIAENVDLFXSWGVTSPEMAPQFVADDTGTFELDSVIQNGYAFADRYDILAMSKNNKYGSK	979
Db	957	RVIYAQNAKLFEKMGITSEFAPQYVSSQDGTFLDSIIENGYAFEDYDIIAMSKNNKYGSL	1016
QY	980	EDLRDALKALHKAGIOAIADWVPDQIYQLPGKEVWTATRTDGAGRIADAIIDHSLYVAN	1039
Db	1017	KOLMDALRALHAEGISAIADWVPDQIYNLPGKEVWTATRTDGAGRIADAIIDHSLYYAAK	1076
QY	1040	SKSSGKDYQAKYGGEGFLAELKAKYPEMKVNMISTGKPIDDSVKLQKWAEBYFNGTNVLE	1099


```
Db 1005 KYGGAFLEELQAKYELPFAKQISTGVPMDFSVKIKQWAKYFNGTNIILRGAGVYLKD 1064
Qy 1109 EATKGYFTVTKGNFIPLQLT---GKEKVIITGFSDDGKIYFTGSGTOAKSAFVTFNG 1164
Db 1065 QATNTYFSLVSDNTFLPKSLVNPNGHTSSSVTGLVPDGGYVYVYSTSGNAKNAFTSLGN 1124
Qy 1165 NTYYFDARGHMTNSEYSPNGKDVVYRFLPGLNGMLSNAPYIDANGNTYL--YNSKGOMYKG 1222
Db 1125 NMYFDNNMYMTGAQ-SINGANY--FLSNGIOLRNAYI--DNGKNVLSYNGDGRRYEN 1180
Qy 1223 GYTKFDVSETDKGSKSVKVFYFTNEGMAGKVTVIDGFQYFGEDGFQAKDLV-TF 1281
Db 1181 GYFLP-----GQQ-----WRYFQ-N-CIMAVGLTRVHGAVQYFDASGFAKQGFITTA 1226
Qy 1282 KKTYYFDHAHTNGIKDTH-RNINGKYYFDANGVAATGAOVINGOKLYFNEDDGGQVKG 1340
Db 1227 DGKLRVDFDRSGNIOISNRNFRVRSKGWFLFDHNGVAVTGTVTFNGQRLYFKFNGVQAKGE 1286
Qy 1341 VVKNADGTYSKYKEGFGELVTFNEFFTDDGNVYVYAGAKTKVTGAOVING 1390
Db 1287 FIRDANGYLRYDDPNSGNEVRNFRVRSKGWFLFDHNGIAVTGARVVG 1336

RESULT 8
T30857
Glucosyltransferase - Streptococcus salivarius
C:Species: Streptococcus salivarius
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30857
R:Simponson, C.L.; Giffard, P.M.; Jacques, N.A.
I:Infect. Immun. 63, 609-621, 1995
A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for pri
A:Reference number: Z20909; PMID:95122197; PMID:7822030
A:Accession: T30857
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1449 <SIM>
A:Cross-references: UNIPROT:Q55264; UNIPARC:UPI00000B166B; EMBL:L35495; NID:G662378; PID
C:Genes:
A:Gene: gtfl

Query Match 38.7%; Score 3067; DB 2; Length 1449;
Best Local Similarity 43.1%; Pred. No. 2.9e-135;
Matches 679; Conservative 204; Mismatches 495; Indels 198; Gaps 38;
Qy 1 MENKHYKLHKVKKQWVTIAVASVALATYLG-GLSVTSSVSADETQDXTV----- 50
Db 1 MDKKVHYKMKVKKQWVTIAVTLGLSLGASVSLGTNDGVQADEHTDATVAIPDITVDT 60
Qy 51 -----TQNSGTTASLVTSPKATKADKRNTEKADVLTPAKET---NAVETATTTT 100
Db 61 GTVSNDDTTAAQDPTTAAVAAATNDVAT---DQATPTATFDLTDTTNTVAANAVIDTVATVGT 117
Qy 101 QATAEATATTADVAAPVNPKEAVVTTDAP--AVTTEK-----AEEQP----- 143
Db 118 DRAATNTDATTADTAVDTNNNTTDTTVDTRAATTBERRATGARGTGGRATPVNG 177
Qy 144 ----ATVKAEVNTEYKAPKAEALKQSEVEAALSLSKNIDKIGYVVYVNEEDSGHKNFAIT 199
Db 178 NTNANNTVTVNNDLPATNNVNTDGP-----SHIKTINGQYVVEDDGTIRKNVYLE 230
Qy 200 VNGQLLYFKED-GALTSSTSYTF--TPGTTNIVDGSIN-----NRAYDSSEASFELI 249
Db 231 RIGGSQYFNAETGELSNOKEYRFDKNGGTGSSADSTNTNTVNGDKNAFYGTGTDKIELV 290
Qy 250 DGLVTDADSYRPAASIKDGVTAQASTAEAFRPLLMAWPNVDVTQVNLNYSKV-FNLDA 308
Db 291 DGYFTANTYRPAELKQKWEWASTENDKRPULLTVWVPSKAIQASLYNMKEQGLGTNQ 350
Qy 309 KYSSTDKQETLVAAKDQIKIEQKIQABKSTOWLRETISAFVQTPOWNKETENYKSG 368
Db 351 TVTSPSSQTMQDAALQEVKRIEERJAREGNTDWRATTKNFVKTPQGNWSENLD--- 407
```

```
Qy 369 GEDHLQGALLVYNDSTRPWNDSYRRILNRTRATNOTGTIDKILDBQSDPNHMGGFDFLL 428
Db 408 NNDHLQGGALLVYNDSTRSHANSDFRLLNRTPTSQTKHNPXYTKDTSN-----GGFEFLU 463
Qy 429 ANDVLSNPVQAEQLNQIHVLMNMGSIUMGDKANFDGIRVDAVDNVDADMLQLYNTYF 488
Db 464 ANDIDNSNPAQAEQLNLHLYIMNIGTTIGSEDEDFDGVVRVDAVDNVNADLQLIASDYF 523
Qy 489 REYGVNKESEANALAHISVLEAWSLNDNHYNDKDTGGAALAMENKORLALLFSLAKPIKR 548
Db 524 KAKYGADQSQAIKHLISLEAWSHNDAYTNEDTKGAQLPMDDPMHLLALVYSLLRPIGNR 583
Qy 549 TPVAPLXNNTFNTTQRDEKTDWINKGSKAYNEDGTVKQSTIGKYNEKYGDASGNVYFI 608
Db 584 S-GVSEPLISNSLN-----DRSESGKNKR-----MANYAFV 613
Qy 609 RAHDNNVQDIIAETIIEKINPKSDGFTITDAEMKQAFIYNDKMLSSDKKYTLNNIPAA 668
Db 614 RAHDSVQSIIGQIIEKINPKSDGFTITDAEMKQAFIYNDKMLSSDKKYTLNNIPAA 673
Qy 669 AVMLQNMETITRYYGDLTYDDGHYMETKSPYYDTIVNLMKSRIKYVSGGQQRSWL-- 726
Db 674 ALMLTHKDTVPRVYGDYTDGQYMAQKSPYDAIETLLKGRIRYAAAGQDMKVNIYGY 733
Qy 727 -PTDGMKDNDSVELYRTNEVYTSVRYGKDIPTANDTEGSKYRSRTSQVTLVANNPKLND 785
Db 734 GNTNG-----WDAAGVLTSVRYGTGANSADT-GTAETRNQGMVIVSNQPALRU- 782
Qy 786 OSAKLNVBMGKIHANQYKRALITVGTADGINKFTSDADAIAAGYVVKETDSNGVLTFCANDI 845
Db 783 -TSNLTINMGAHRNQAYRPLLLTTNDGVATVLYDSD--ANGIVKYTDGNGNLTFSANEI 839
Qy 846 KYETFDMSGFVAVWVPVVGASDNQDIRVAPSTPEAKKEGELTLUKATEAVDSQLIYGFSNF 905
Db 840 RGIERNPQDGYLAVWVPVVGASENQDVRVAPSKKXSSG-LVYESNAALDSQVIYEGFSNF 898
Qy 906 QTIPOGSDPSVYTNKIAENVDLFKSMGVTSPKAPQFVSADDTGLFSDSVIONGVAFADR 965
Db 899 QDFV--QNPQSYTNKIAENANLKFSGWGITSEFAPQVYVSSDDGSFLDSVIONGVAFDR 956
Qy 966 YDLAMSNNKYGSKEDLRDALKAHKAGIQATADWVPQIYQLPGKEVVVTAATRTDAGRK 1025
Db 957 YDIGNSKDNKYGLADLKAALKSLHAGVLSADWVPQIYNLPGDEVVTAATRVNNGYT 1016
Qy 1026 IADAIIDSLYVANSKSGDYQAKYGGFEFLAELKAKYKPEMPKVMNISTGKPIDDSVKUK 1085
Db 1017 KDGAIIDHSLVAAKTRTFGNDYQGYGAFLDELKRLYPQIFDRVQISTGKMTTDEKIT 1076
Qy 1086 QWKAQYFNGTNVLERGVYVLSDEATKYFTVTKESGNFPL-OLTKCKEVIITGFSDDGK 1144
Db 1077 QMSAKYMGNTNIDRGSEYVNLKNGLNGYYGT---NGKRVSLPKVVGSNOSTNGDNQNGDG 1133
Qy 1145 -----ITYFGTSGTQAKSAFVTFN-GNTYFFDARGHMTNSENSEYSPNGKDVYRFLP 1193
Db 1134 SGKFEKRLFSVRYRYNNGQYAKNAFIKNDNGVYVYFDNSGRMAV-GEKTIQDKQ-YFFELA 1191
Qy 1194 NGIMLSNAPYIDANGNTYLNSKGOMYKGYTKFVDSQTKDQKESKVVKFRYFTNEGVM 1253
Db 1192 NGVQLRDGVQRNRRQGVFYDQNGVLNANG-----KQDPKP----- 1227
Qy 1254 AKGVTVIDGTYQYFGEDGFQAKDLVTFKGYTYYFDAHTNGIYKDTWRNNGKVVYFDAN 1313
Db 1228 -----DNNNNASGRNQF-----VQIGNNV-----WAYYDGN 1253
Qy 1314 GVAATGAOVINGOKLYFNEEDSGQVKGVVKNADGTYSKYKEGFGELVTFNEFFTDDGNVWY 1373
Db 1254 GKRVTGHQNINGQELFPDNNGVQVKGRTV-NENGAIYYDANSGEWARRNFAEIEPQVNA 1312
Qy 1374 YAGANGTKVTGAOVINGOHLYPNADGSQVKGVVKNADGTYSKYKNASTGERLUTNEFFTGG 1433
Db 1313 YFNNDGTAVKGSQNINGQDLVFDQNGRQVKGALA-NVDGNLRYDYVNSGELYRNRPHET- 1370
Qy 1434 DNNWYIYGANGKSVTGEVKIGDDTYFFAKDGGKQVKGQTVSAGNGRISYYYYGSGKRAVST 1493
```


[illegible]

[illegible]

Qy 1447 VTGEVKIGDDTYFFPAK-DGKQVKGQTVSAGNGRISYYYGDSG 1487
Db : : : : : : : : : : : : : : : : : :
Dg VKGFKVNNKLQHFEITGVTKSAHIIIVNRR--TYIFDDQG 1502

RESULT 13
B97033
uncharacterized protein, related to enterotoxins of other Clostridiales [imported] - Clc
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004
C:Accession: B97033
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.: Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clc
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: B97033
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2817 <RUR>
A:Cross-references: UNIPROT:Q97K42; UNIPARC:UPI00000CA0A0; GB:AE001437; PIDN:AAK79053.1.1
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC1079

	Query Match	7.3%;	Score 582.5;	DB 2;	Length 2817;
	Best Local Similarity	22.3%;	Pred. No. 8.3e-19;		
	Matches 382;	Conservative 179;	Mismatches 533;	Indels 621;	Gaps 96;

```

Qy      36 TTSSVASDETOKTVTQSNGTTASLVTSPETAKEDKRTNTKEADVLTPAKETNAVETA 95
Db      ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy     153 STSSAQTKGSNDNNIPSNNTSTNTSKNEPSNTD-----IKTTEAPANAPIKD----- 200
Db      ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

Qy     96 TTTNTQATAEAAATTATADVAVAAPNKBAVVTT---DAPAVTTEKAEEOPATVKAAYVN 152
Db      ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Dg    201 -TPNQSDSALAKNALSNLNLAADSQTSKVTSSNDA PKVNTTSTDK-----KASNLN 254

Qy    153 TEVKAPEALKDSEVEEAALS LKNINIDG-KYYIYNEDGSHKENFALT VNGQLYFGKDG 211
Db      ||| ||||| : : | : | : | : | : | : | : | : | : | : | : | : | :
Dg   255 ND-----SQDGWTK-----DGKKYYIYN--GVQQKGFO-SINKSIYFNDDG 294

Qy    212 ALT-----SSSYTGFTPGTTNVDFGSINNRAVDSEASFELIDGYLTADSWRPASII 265
Db      ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Dg   295 SMQTGLKYNSNYFYFDASGVMLTGLQNIGTYGYFNDDG-KLLTGLQAINNNYYFN-- 351

Qy    266 KGVTWQAStADBFRPLLMAMPNVDTQVNYLN---YMSKFNLDAKYSSTDQETLKVA 322
Db      ||| ||||| : : | : | : | : | : | : | : | : | : | : | : | : | :
Dg   352 NDTV-----MQGTWTCNDSKYFYFDNNGVMOTGLVHNKNNKYGFNGDKLLTG 399

Qy    323 AKDIQIKIEQIAEKSTOWLBETISAFVKYTOPWNKETENYSGGGEDHLGGALLIYN 382
Db      ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Dg   400 LQNI-----NNYT-----YYF 410

Qy    383 DS----RTPWANSD---YRLNRATATNOTGITDKSILDEQSDPNHMGGFDLLANDVDL 434
Db      ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Dg   411 DSVGWMQTDWTIDISKYFYSVN--GMQTGI-----YISGYIYGFANDGKL 456

Qy    435 SN--PVQOAE----QLNQIHLYMWSGI-----VMGDKANPD-----GIR 469
Db      ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Dg   457 LTGLQVINGNSYFYFDNGIRLVSRWITIDGKYVFYFNQDGILTDNWINVDGKYFYISGVK 516

Qy    470 VDADVNDVADMQLYNTNYFREY-----GVNKSEA-----NALAHLSVLEAW-SL 513
Db      ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Dg   517 QTGLQNIDG-----NRY--YFDSSGIMQTLQKIDGKTYTFDNGIRQI-----GWITY 563

Qy    514 NDNHYNDRKTGGAALAMENQRLLALSLAKPIKERTPAVSPVYNNTF----- 560
Db      ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Dg   564 QNNKYFNSDG-----SMQTDLIKYSYSTSP-YNYHYQYYGFDNDGKLLT 607

Qy    561 -----NTQRD-----EKTDMINKGSKAYNEDGTWKQSTIGKYNKYGASGNVYFIR 609
Db      ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Dg   608 GLQTIKGNYYFPDSNGISOMGWVINIDGKDPFYNSNSIMTENWVINDEKY-----YFYI- 660

```

Qy	610	AHDNNVQDIIAEI	IKKEINPKSDGTTI	DAEHKQAFBI	YNKDMLSDD	KKYTLNNIP	AAAYA	669					
Db	661	---NNV---	---KQTGYINGK---	---YYFDFD---	---	---	---	682					
Qy	670	VNLQNMETIT	-RVYYGDL	YTDGHHYMETKSPY	YDITIVNLMSRTIKYVSGQ	AQBSYMLPT	728						
Db	683	IMQTGFQIISGNTYYLD	---DNG---	-VKQTGWVTI---	-KGQYIFD	GNQVMINYYW---	729						
Qy	729	DGKMDNSDVELYRTNEVYTSV	-----RYGKD	IMTANDTEGSKY	SRTSGQVTLVAN	PKL	782						
Db	730	---FNDKNTYYVINGNM	QTGAISINNHHYGF	---DNGIM	QTGMQRINGR	TYYPDNN---	780						
Qy	783	NLDQSAK	-LNVEMG	KIHANOKYRALIVGT	ADGINKNFTSDADA	IAAGVVKETSDSNG	UTF	840					
Db	781	---GAAKTGLVTEGKTY	FFNTYAYL---	---	---	---	811						
Qy	841	GANDIKGYETFDMSGF	VAV-WPVGAS	---DNQD	IRVAPSTEAKG	GELTLKATE	AYS	895					
Db	812	NNN	---YF	LNNGVVTRTGINTSYNN	RYLDJSTGVRV	---	845						
Qy	896	QLIYEGFSNFOTI	IPDGDPSVYTR	NKRIKAENVDL	LFKSGVTSFEMA	POFVSADDTFL	DSV	955					
Db	846	-----TGFTI	-DG-----NKYY	FDSSCA---MCT	SPIIV---	---	871						
Qy	956	IQNGYAFADR	YDLAWSKNNKYG	SKEDJRDALKALH	KAGIAQIADW	PQIYQLP	KEVYT	1015					
Db	872	--NG--	---NTYG	FSKD-----GI---	---	MLTGW	QITL	892					
Qy	1016	ATRIDGAGR	KIADAIIDHSL	VYANSKSS-----GKDY	-QAKYG	GEFLA-ELKAK	1062						
Db	893	SSNTYS	-----SYNI	YFNSDGSQAQGF	FTYLGKTYFP	FPNTGYMLL	GYNYINGK	941					
Qy	1063	YPENFKVNM	I	STGPI	DDSVKVK-----QWKA	EYENGTVNL	ERG	VYVLS	1108				
Db	942	YYYFDNG	VIQTGWTR	DRSSKY	LDPSGA	AVTGFQNTINGDKY	YFNSGIM	QTLGV	YNPD	1001			
Qy	1109	EATGKYFTV	TKEGNFIPLQ	LTGKEKV-----ITG	FSDDGKG---ITY	FCTS---GTOAKSA	1158						
Db	1002	-----YYG	FDNDNGHI---	-LTGMS	INGYIYYFD	STGKAQGFV	TYLGKTYFYFN	TYVTG	1052				
Qy	1159	PVTNGNTY	FPDARGHMT	-----NSE	---YSPNG	KDV-----YRPL	PN	GIMLSN	1200				
Db	1053	FVNANNLLY	FPDNEGVMTG	WINTNSNRY	YFSATGAS	VTGQTIDG	NKYCF	DSNG	1107				
Qy	1201	AFYIDA	--NGNT	YLXNSKGM-----YKGY	TK-----FDV	SETDKDG	ESK	VWK	1243				
Db	1108	AIYD	VVTINGSTYGF	NTDGLMLTGM	QTI	RYNRYGSSYFN	TYFNSD	GTAKTG	FPTVLYNK	1167			
Qy	1244	FRYPT	-NEGVM	AKVTVIDG	FTQYFG	EDGFGF-----AKDK	LV	TFKG	1283				
Db	1168	TYFNP	SDGRMLQGYO	YINGNHYYFAD	GTMTQ	GWITNGSSKY	YLDPSGA	AVTGLQ	TING	1227			
Qy	1284	KTYFY	FDH---	TGNGIK	DTWRN	INGKYYFD	ANGVAATGAQV	INGKYL	FNE	DGSO	VKG	1340	
Db	1228	NKYCF	DSNGI	LQHN	GI-----FYIGN	TYGSDNN	GIMLTGLQ	LINGLY	CFNS	DSGS-VKTG	1282		
Qy	1341	VVKNA	DQTY--SKY	KG	FGE	LVTE--FF	TTDGNV-----W	---Y	GANGK	TVTGAQ	1386		
Db	1283	LVTYL	GKTYFY	DFSYSVSG	FQINNN	TYFGNDG	MTQMGWNYGY	RYL	YLD	SGIKV	TGWQ	1342	
Qy	1387	VINGO	HLVFNADG	SOVKG	VKNADG	YVSKYNAST	GERLTNEP	TTTCD	NN	WYI	GANGKS	1446	
Db	1343	TIDG	KYFYDYGA--KTG	IV-NIDG	NYGFN-NSG	WMLTG	WQHING	ST--YF	NSNGIA	1396			
Qy	1447	VTGE	VKIGD	DDTYFFAK	DGK-QV	GQ	TVSAGN	GRISYY	YIG	SGKRAV	ST-----WIE	1496	
Db	1397	NTG	FTI	VLGKTY	FDSDYSGR	MQIGSM	TI--NG-TS	YFYANG	VMK	TTSD	SP	TLAVCWVR	1452
Qy	1497	-----	-----IQB	VYV	FDK	KNGL	1510	---	---	---	---	---	
Db	1453	DSYYYQ	YVYLN	AAAGTK	LTGLQ	TIDGNTY-YF	DSNGI	1486	---	---	---	---	

Db 732 -----QDSITVSANOYEVRIINEBGRILD---HSGRWINKESIIKD-----ISSK 775
Qy 277 E--DPRPLMAMPNVNDTVNYLYNMSKV-F-NLDAKYSTTDKQOETLKVAARDIQIKIEQK 333
Db 776 EYISFNPK-----ENKIIKSVKYLHELSTLLQEIERNANSSDIDLEKKVWLTECEINVASN 831
Qy 334 IQAEKSTOWLRETISAFVKTOPOWNKETENYSKGGEDHLOGGALLYVNDSTRPWNASDY 393
Db 832 ID-----RQIVGREIEBAKNLTSDSINYIK-----NEF 859
Qy 394 RRLNRTATNQTGIDKISILDEQSDPNHMGFD---FLANDV-----DLSN 436
Db 860 KLI-----ESISLSYDLKHONGLDUSHFISFEDISKTEFNGFRIRINKETGN 907
Qy 437 PV-VQAEQLNQIHYLMNWSIVMGDKDANPDG-----IRVDAVNDVADMLQLYTNY 487
Db 908 SIFETEKEITFSEYATHISKEISNIKOTIFDNVNGKLVKKVNLDAHAHEVNTLSAFFIQS 967
Qy 488 FREYGVNKSSEANALAHISVLEAWSLNDNHNDKTDGAALA-----ME 530
Db 968 LIBYNTTKESLSNLSVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETIDLLPTLSE 1027
Qy 531 NKQRLALLF---SLAKPIKERTPAVSPLYNNTENTTQRDEKTDWINKGSKAYNEDGTVK 587
Db 1028 GLPIIATIIDGVSUGRAIKELSETNDPLLRQEIEA-----KIGIMAVN-----LT 1072
Qy 588 QSTIGKYNEKYGDASGNYPF-----IRAHDNN---VQDIIABIIKKEINPKSDGF 634
Db 1073 AASTAIVTSALGIASGSIILLVPLAGISAGIPSLVNNELIQLDKATKVI-----DYF 1124
Qy 635 T-ITDAEMQAPFIYNKMDLSSDKYTLNIPPAAYVMLQNMETIT---RVY-----VG 684
Db 1125 KHISLAETGAFTLLDDKIIMPQODLVLSBIDF-----NNSITLGRKEIWRABEGSG 1177
Qy 685 DLVTD-DGHMETKS-----PY---YDTIVNLMKSRIK-----VSGGQAQRSY-----W 725
Db 1178 HLTLDIDHFFSPSYTRKPLWSIYD-VLNIIKEKIDFSKOLMVLDPNAPNRFVGYEMGW 1236
Qy 726 LPTDGRMDNSVELY-RTNEVY-----TSVRYGKD-----754
Db 1237 TPGFRSLDNGTKLLDIRIDHYEQFYRYFAFIADALITKLKPRVEDINVLNDGNTR 1296
Qy 755 -----IMTANDTEGSKYS-RTSGQVTLVANNPKLNLDQS-----787
Db 1297 SFIVPVITTEQIRKRLNSFYSGSGSYSLSPYNNNIDLNLVENDTWVIDVNVVKNIT 1356
Qy 788 -----AKLNVE-----MGKIHAKQYBALIVGTADGKNFT 818
Db 1357 IESEDEIQGELIENILSKLNIENDKIIILNNHTINFYGDINESNRFISLTFISLEDI-NII 1415
Qy 819 SDADAIAGYVKETDSNGV-LTFGANDIKGYETFDMSGFVA---VMVPVGASDNODIRVA 874
Db 1416 IEIDLVSYSKYILLSCNCKLIENSSDIO-QKIDHIFNGEHQKYPISYIDNETKYNG 1473
Qy 875 PSTEAKEGELTLKATEAVDSQLIYVEGFNFOTIPDGSDPSVYTNR-----KIAENV-D- 927
Db 1474 FIDYSKKEGLFTAESFNSIESIIRNIY-----MPDSNNLFTYSSKDLKDIRIINKGDVK 1525
Qy 928 -LFSKNGVTSFEMAPQFVSAD-----DGTFLDSVIQNGYAFADRVDLAMSNNKYSGKE 980
Db 1526 LLIGNYFKDDMKVSLSTIEDTNTIKLNGVYLD---ENGVA-----QILKFMNNAKSALN 1577
Qy 981 DLRDALKALHKAGIOAI--ADWVPQIYQLPGKEVVVTRTDGAGRKIAD-----1028
Db 1578 TSNLNMNFIENIKNI FYNLDPNIEFLDTNFIISGNSISGOPELICDKDKNTQPYFI 1637
Qy 1029 -----AIIDHSLYVANSKS-----SGK-----DYQAK--YGGEFLAELKAKYPE 1065
Db 1638 NEFIKETSITLYVGNRQNLIVEPSVHLDSDGNSISSTVINFSQKLYGIDRYVNVKVIAPN 1697
Qy 1066 MF--KVNMLSTGKPIDDSVKLQWKAIEYFN---GTNVLERGQVYLSDEATG-----1112
Db 1698 LYTDEINITPVYKPNYICFEVILDANYINEKINVNINDLSIRYVNDNDGSLILIANSE 1757

Qy 1113 -----KYFTVTKEG-----NFIPLOLTGKEKVTGFS-----SDG-----1142
Db 1758 EDNQPOVKIRFVNVFSDTAADKLSFNFSDKQDVSVSKIIISTSLAAYSOGFPDYEGLV 1817
Qy 1143 -----KGITYFOTSGTOAK-----SAPVTNGNTYYFD-ARGHMVTN 1178
Db 1818 SLNDYFYINSFGNMVSGLIYIINDSLYFPPKNNLITGTFTTIDGNKYYPDPTKSGAASI 1877
Qy 1179 SEYSPNGKDVY-----RFLPNGIMLSN-----AF 1202
Db 1878 GEITIDGKYYFNKQILQVGVINTSDGLKYFAPAGTLDENLEGESVNFIGKLNIDGKIY 1937
Qy 1203 YIDAN-----GNTYLYNSK-GOMYKG-----GYTK--PDVSETDKDGKESKVVKF 1244
Db 1938 YPEDNRAAAVEWKLDDETYYFNPKTGEALKGLHQIGDNKYYFDDNGIMGTGITINDKV 1997
Qy 1245 RYFTNEGVMAKVTVVIDGFTQYFGDG-----POAKDK-----1277
Db 1998 FYFNNDGVNQVGYIEVNGKYFYFGKNGERQLGVFNTPDGFKFPKDDDLGTEGELTY 2057
Qy 1278 --LVTFPKGTYTFD---AHTNGIKDTRWNINGKWWYFDAN-GVAATGAQVINGOKLYF 1330
Db 2058 NGLNENFKIYTFDISNTAVVGWGLD-----DGSTYYFDDNRAEACIGLTVINDCKYF 2112
Qy 1331 NEDGSQVKGWVKNADGTYSKYKEGFGELVLTNEFTTIDGNVWYVYAGANGKTVTGAOVING 1390
Db 2113 DDNGIRQLG-----FITINDNIFYFS-ESGKIELGYQNING 2147
Qy 1391 QHLYFNADGSQVKGWVKNADGTYSKYNASTGERLTNEFTTIDGNWYVYIANGKSV--T 1448
Db 2148 NYFYIDESG-LVLIGVFDTPDG--YKYFAPLN-----TVNDNIY-----GOAVKYS 2190
Qy 1449 GEVKIGDDTYFPAKDGKQVKGOTVSAGNCRISYYVGDSCBRAVSTWIEIOPGVVYVFDKN 1508
Db 2191 GLVRNEDVYTF---GETYKIETGIENETDKTYFDPETKKAYK-GINVVDDIKIYFEDEN 2246
Qy 1509 GL 1510
Db 2247 GI 2248

Search completed: February 11, 2006, 19:42:57
Job time : 44.3322 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.
OM protein - protein search, using sw model
Run on: February 11, 2006, 18:57:35 ; Search time 204.83 Seconds
(without alignments)
5228.676 Million cell updates/sec
Title: US-10-797-821-40
Perfect score: 7928
Sequence: 1 MENKHYLHKVKKQWNTIA.....PGVYVYFDKGLAYPRVLN 1518

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 2166443 seqs, 705528306 residues
Total number of hits satisfying chosen parameters: 2166443
Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : Uniprot_05.80.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES										
Result No.	Score	Match	Length	ID	Description					
1	7928	100.0	1518	2	Q00600_STRSL	Q00600	streptococc			
2	3973.5	50.1	1512	2	Q9WJ5_9STRE	Q9WJ5	streptococc			
3	3949.5	49.8	1506	2	Q56CX8_9STRE	Q56CX8	streptococc			
4	3818	48.2	1554	2	Q8KZL5_9STRE	Q8KZL5	streptococc			
5	3743	47.2	1599	2	Q00599_STRSL	Q00599	streptococc			
6	3532	44.6	1462	1	GTFF_STRMU	P49331	streptococc			
7	3373.5	42.6	1577	2	Q54178_STRGN	Q54178	streptococc			
8	3350.5	42.3	1575	2	Q9LCH3_STROR	Q9LCH3	streptococc			
9	3325.5	41.9	1590	2	Q59983_9STRE	Q59983	streptococc			
10	3304.5	41.7	1597	1	GTFL_STRDO	P11001	streptococc			
11	3302.5	41.7	1590	2	Q5263_9STRE	Q5263	streptococc			
12	3296.5	41.6	1592	1	GTFF_STRDO	P27470	streptococc			
13	3239	40.9	1476	1	GTFF_STRMU	P08987	streptococc			
14	3183.5	40.12	1455	1	GTFC_STRMU	P13470	streptococc			
15	3155	39.8	1577	2	Q5265_STRSL	Q5265	streptococc			
16	3067	38.7	1449	2	Q5264_STRSL	Q5264	streptococc			
17	3056	38.5	1449	2	Q68542_STRSL	Q68542	streptococc			
18	3053.5	38.5	1454	2	Q69A94_LEUME	Q69A94	leuconostoc			
19	2883	36.4	2835	2	Q8G9Q2_LEUME	Q8G9Q2	leuconostoc			
20	2840.5	35.8	1338	2	Q9WJ4_9STRE	Q9WJ4	streptococc			
21	2804	35.4	1365	1	GTFS_STRDO	P29336	streptococc			
22	2794.5	35.2	1561	2	Q58M8_9LACO	Q58M8	lactobacill			
23	2768.5	34.9	1522	2	Q6TXV4_LEUME	Q6TXV4	leuconostoc			
24	2759.5	34.8	1527	2	Q8KRE1_LEUME	Q8KRE1	leuconostoc			
25	2726.5	34.4	1527	2	Q9ZAR4_LEUME	Q9ZAR4	leuconostoc			
26	2721	34.3	1290	2	Q48756_LEUME	Q48756	leuconostoc			
27	2638.5	33.3	1508	2	Q52224_LEUME	Q52224	leuconostoc			
28	2627.5	33.1	1508	2	Q9EZH5_LEUME	Q9EZH5	leuconostoc			
29	2599.5	32.8	1477	2	Q91466_LEUME	Q91466	leuconostoc			
30	2557.5	32.3	1595	2	Q58M3_LACSK	Q58M3	lactobacill			
31	2543.5	32.1	1330	2	Q84CN4_LEUME	Q84CN4	leuconostoc			

RESULT 1
Q00600_STRSL PRELIMINARY; PRT; 1518 AA.
AC Q00600;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DE 01-FEB-2005 (Tremblrel. 29, Last annotation update)
DE Glucosyltransferase-I.
GN Name=gtfJ;
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 25975;
RX MEDLINE=92148377; PubMed=1838391;
RA Giffard P.M., Simpson C.B., Milward C.P., Jacques N.A.;
RT "Molecular Characterization of a cluster of at least tow
RT glucosyltransferase genes in Streptococcus salivarius ATCC 25975.";
RL J. Gen. Microbiol. 137:2577-2593(1991).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 25975;
RA Jacques N.;
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z11873; CAA77900.1; -; Genomic_DNA.
DR EMBL; M64111; AAA26896.1; -; Genomic_DNA.
DR FIR; A44811; A44811.
DR HSP; P06653; LGVN.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW binding_1; 2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
DR Transference.
SQ SEQUENCE 1518 AA; 167730 MW; DAA41F717098B59A CRC64;
Query Match 100.0%; Score 7928; DB 2; Length 1518;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1518; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MENKHYLHKVKKQWNTIAVASVALATVLGLSVTTSSVSADETQDKVTQNSGTTAS 60
DB 1 MENKHYLHKVKKQWNTIAVASVALATVLGLSVTTSSVSADETQDKVTQNSGTTAS 60
QY 61 LVTSPEATKEADKRTNKTEADVLTPAKETNAVETATTTNTQATAEAATTTATTADVAVA 120
DB 61 LVTSPEATKEADKRTNKTEADVLTPAKETNAVETATTTNTQATAEAATTTATTADVAVA 120
QY 121 PNKEAVTTTADPAVTTKEAEQOPATVKAENVTEVKAPAAALKDSEVEAALSUKNTKID 180
DB 121 PNKEAVTTTADPAVTTKEAEQOPATVKAENVTEVKAPAAALKDSEVEAALSUKNTKID 180

ALIGNMENTS

QY 181 GRKYVYVNEBDSGSHENFAITVNGQLLYFGKDGALTSSTYSFPTGTTNIIVDGSFINNAYD 240
DB 181 GRKYVYVNEBDSGSHENFAITVNGQLLYFGKDGALTSSTYSFPTGTTNIIVDGSFINNAYD 240
QY 241 SSEASFELIDGVLTDADSWRPAASIIKDGVTWQASTAEDFRPLLMWPNVDQVNLNYM 300
DB 241 SSEASFELIDGVLTDADSWRPAASIIKDGVTWQASTAEDFRPLLMWPNVDQVNLNYM 300
QY 301 SKVFNLDAKYSYTDQETLKVAAKDIQIKIEQIKOAKESQWLRRTISAFVKTQOWNKE 360
DB 301 SKVFNLDAKYSYTDQETLKVAAKDIQIKIEQIKOAKESQWLRRTISAFVKTQOWNKE 360
QY 361 TENYKGGGDEHLOGGALLVYNDSTRTPWANSYRRLNRATNTQGTIDKSIILDEQSDPNH 420
DB 361 TENYKGGGDEHLOGGALLVYNDSTRTPWANSYRRLNRATNTQGTIDKSIILDEQSDPNH 420
QY 421 MGGFDFLLANDVDLGNPVVQASQLNQIHYLMNWGSTVMGDKDANPDGIRVDADVNDADM 480
DB 421 MGGFDFLLANDVDLGNPVVQASQLNQIHYLMNWGSTVMGDKDANPDGIRVDADVNDADM 480
QY 481 LQLYTNYFREYYGVNKSSEANALAHISVLEAWSLNDNHNDKTDGAALAMENKQRLALLFS 540
DB 481 LQLYTNYFREYYGVNKSSEANALAHISVLEAWSLNDNHNDKTDGAALAMENKQRLALLFS 540
QY 541 LAKPIKERTPAVSPLYNNTFNTQREKTDWINKOGSKAYNEBGTGTVKOSTIKYNEKYGD 600
DB 541 LAKPIKERTPAVSPLYNNTFNTQREKTDWINKOGSKAYNEBGTGTVKOSTIKYNEKYGD 600
QY 601 ASGNVYFTRAHONNVQDIIAEIIKKEINPKSGDFTITDAEMKOAFEIYNKDWMLSDKKYT 660
DB 601 ASGNVYFTRAHONNVQDIIAEIIKKEINPKSGDFTITDAEMKOAFEIYNKDWMLSDKKYT 660
QY 661 LNNIPAAVAVMLQNMETITRVYVGDLYTDDGHYMETKSPYYDTIIVNLMKSRIKYVSGQA 720
DB 661 LNNIPAAVAVMLQNMETITRVYVGDLYTDDGHYMETKSPYYDTIIVNLMKSRIKYVSGQA 720
QY 721 QRSYWLPTDGKMDNSDELRYRNEVYTSVRYGKDIIMTANDTEGSKYSRSTSGQVTLVANNP 780
DB 721 QRSYWLPTDGKMDNSDELRYRNEVYTSVRYGKDIIMTANDTEGSKYSRSTSGQVTLVANNP 780
QY 781 KNLNQSACLNVEMGKIHANOKYRALIVGTADGINKFTSDADAIAAGYVYKETSNGVLTF 840
DB 781 KNLNQSACLNVEMGKIHANOKYRALIVGTADGINKFTSDADAIAAGYVYKETSNGVLTF 840
QY 841 GANDIKGYETFDMSGFVAVWVPVGASDNQDIRVAPSTEAKKEGELTLKATEAYDSQLIYE 900
DB 841 GANDIKGYETFDMSGFVAVWVPVGASDNQDIRVAPSTEAKKEGELTLKATEAYDSQLIYE 900
QY 901 GFSNFQTTIPDGSDPSVYVTRKIAENVDLPKSGWGVTSFEMAPQFVSADGTFDLSVIQNGY 960
DB 901 GFSNFQTTIPDGSDPSVYVTRKIAENVDLPKSGWGVTSFEMAPQFVSADGTFDLSVIQNGY 960
QY 961 AFADRYDLAWSKNNYKSGKEDIRDALKALHKAGIQAIADWVPDQIYOLPGKEVVTATRD 1020
DB 961 AFADRYDLAWSKNNYKSGKEDIRDALKALHKAGIQAIADWVPDQIYOLPGKEVVTATRD 1020
QY 1021 GAGRIADAIIDHSLYLVANSKSGDYQAKYGEFELAEIKAKYPEMFKNMISTGKPIDD 1080
DB 1021 GAGRIADAIIDHSLYLVANSKSGDYQAKYGEFELAEIKAKYPEMFKNMISTGKPIDD 1080
QY 1081 SVKLQWKABEYFNGTNVLERGVGYVLSDEATGKYFTVTKEGNFIPLQLTGKSKVITGFSS 1140
DB 1081 SVKLQWKABEYFNGTNVLERGVGYVLSDEATGKYFTVTKEGNFIPLQLTGKSKVITGFSS 1140
QY 1141 DGKGIYFCTGTOAKSAFVTNGNTYFEDAGHVMVTNSEYSPNGKDVYRFLPNGIMLSN 1200
DB 1141 DGKGIYFCTGTOAKSAFVTNGNTYFEDAGHVMVTNSEYSPNGKDVYRFLPNGIMLSN 1200
QY 1201 AFYIDANGNTYLYNSKGQYKGYTKFDVSETDKDGKESKVVKFRYFTNEGVMAGVTVI 1260
DB 1201 AFYIDANGNTYLYNSKGQYKGYTKFDVSETDKDGKESKVVKFRYFTNEGVMAGVTVI 1260

QY 1261 DGTQYFGEDEGFOAKDKLVTRFKGTYYPDAHTNGIKDTRWNINGKMYYPDANGVAATGA 1320
DB 1261 DGTQYFGEDEGFOAKDKLVTRFKGTYYPDAHTNGIKDTRWNINGKMYYPDANGVAATGA 1320
QY 1321 QVINGQKLYFNEBDSQVKGVVKNADGTYSKYKEGFELVTNEFFTIDGNVWYVYAGANGK 1380
DB 1321 QVINGQKLYFNEBDSQVKGVVKNADGTYSKYKEGFELVTNEFFTIDGNVWYVYAGANGK 1380
QY 1381 TVTGAQVINGQHLYPNADSGQVKGVVKNADGTYSKYNASTGERLTNEFFTIDGNNNWYI 1440
DB 1381 TVTGAQVINGQHLYPNADSGQVKGVVKNADGTYSKYNASTGERLTNEFFTIDGNNNWYI 1440
QY 1441 GANGKSVTGEVKIGDDTYFFAKDKQVKGQVTSAGNGRISYYYGSGKRAVSTWIEIQBG 1500
DB 1441 GANGKSVTGEVKIGDDTYFFAKDKQVKGQVTSAGNGRISYYYGSGKRAVSTWIEIQBG 1500
QY 1501 VVYVFDKNGLAYPPRVLN 1518
DB 1501 VVYVFDKNGLAYPPRVLN 1518
RESULT 2
Q9WJ5_9STRE PRELIMINARY; PRT; 1512 AA.
AC Q9WJ5_9STRE PRELIMINARY; PRT; 1512 AA.
DT 01-NOV-1999 (TremBLrel. 12, Created)
DT 01-NOV-1999 (TremBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)
DE GTP-S.
GN Name=gTfT;
OS Streptococcus criceti.
OG Plasmid pAMI.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1333;
RN [1]_TaxID=1333;
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=HS-6;
RA Inoue M., Fukui K., Miyagi A.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB026123; BAA77237.1; -; Genomic_DNA.
DR HSSP; P06654; 1MPE.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR Pfam; PF01473; CW_binding_1; 2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
DR Plasmid.
SQ SEQUENCE 1512 AA; 167145 MW; 4C03D9C8C601FC14 CRC64;
Query Match 50.1%; Score 3973.5; DB 2; Length 1512;
Best Local Similarity 52.9%; Pred. No. 2e-170;
Matches 811; Conservative 216; Mismatches 415; Indels 91; Gaps 24;
QY 1 MENKHYHLKVKKQWVTIAVASVALATVGLSVTTSVSSADETQDKTVTOSNCSGTAS 60
DB 1 MERKLYKLKVKQWVTIAVASGLASVIGA-GAASQTVSADDICNGASASAEQNTSAS 59
QY 61 L---VTSPEATKEADKRTNTKEADY-LTPAKETNAVET--ATTNTQTAAATATTATTA 113
DB 60 QNKEVVDSDAQAQATDAKATSEQAASVSDVTNSETDQVQNVGDVGSANNQPEAPAAQAAAS 119
QY 114 ---DVAAVAPNKEAVVTTDDAPVTTTEKAEQOPATVKAEEVNVTEVKAPBAALKDSEVAA 170
DB 120 NNTNTATSEANTNTAVSEAAAPAAENRTAEK-----GADLSQDEAAAA 162
QY 171 LSLKNIKNIDGKYVYVNEBDSGSHENFAITVNGQLLYF-GKDGALTSSTYSFPTGTTNIV 229
DB 163 LSLDNIKKVDGKYVYVNMADGSYKKQFAITVKGQLLYFDAETGALSSTYSFSQGLTPLV 222
QY 230 DGFSINNRAYDSSEASFELIDGLTADSWYRPAASIIKDGVTWQASTAEDFRPLLMWPN 289
DB 223 SDFSINNKAFFDSAKSFLVDGLYLTAEWSYRFTKILENGKTWVDSKETDLRPVLTSMWPD 282

```
QY 290 VDTQVNLNTMSKVFNLDAKYSSTDKQETLKVAADQI QIKIEQKIQAEKSTOWLRITISA 349
Db 283 KDTQVAYLNTMSKALGCKEBFTTKTSQTALNTAAEMI QMKIEQRISEKQCTAMLRDAMAA 342
QY 350 FVKTQPMNKETENYSKGGEDHLOGGALLYVNDSTRTPWANSYRRLNRTATNQTGTIDK 409
Db 343 FVATQSRWNSDSQDPK---NDHLOGGALLYTNKLTEWADSRYRLNRTPTTQDGKTHY 399
QY 410 SILDEQSDPNHMGGFPLLANDVDLSNPVVOAQLNQIHYLMWGSIVMGDKDANFDGIR 469
Db 400 SKADE-----YGYEFLANDVNSNPVVOAQLNQIHYLMWGSIVMGDKDANFDGIR 453
QY 470 VDAVDNVDMLQLTYNYFREYYGVNKSSEANLAHISVLEAWSLNDNHNDKTDGAALAM 529
Db 454 VDAVDNVDADTLQLTYNYFNNAVYGVDSKSEAQALAHISILEAWSYNDNYNQDINGAALAM 513
QY 530 ENKQRLALLPSLAKPIKERTPAVSPLYNNFTNTQRBKTDWINKDGSKAYNEDGTGKOS 589
Db 514 DNGLRSLSLYTLRPLSERTPGLSTLIKSEYGLTDRTK----- 551
QY 590 TIGKYEKYGDSAGNVYFIRAHNNVQDIIAEIIKKEINPKSDGFTITDAEMKQAFEIYN 649
Db 552 ----DQKYGDTQPSYFVRAHDSVQTVIAQIIKEKIDPTDGTFTLDQLKQAFDIYN 606
QY 650 KQMLSSDKKYTLNMIIPAAVAVMLQNNMETITRVVYGLYTDGHHYMETKSPYYDTIIVNLMK 709
Db 607 KQMNVSVEKHYTHYNIIPAAVAVMLSNMESVTRVYVYGLYTDGHHYMETKSPYYDAINTLRL 666
QY 710 SRIKYVSGGQA---QRSYMLPTDGMQNSDV-ELYRTRNEVYTSVRYKQDIWANTDTEGSKY 766
Db 667 ARIRYAAGGQTMHEKAYTPSAAMAKANPPDSGSLGNSEVLVSVRFGQDVMSADMTGGQL 726
QY 767 SRTSGQVTLVANNPKLNLDQSAKLNVEMGKIHANOKYRALIVGTADGINKFTSDADAIAA 826
Db 727 AKTSGMFLSIANNPELELDANEIEIKVNVGKIHAQYRPLLLTDDKGLQKYLNDSD---T 783
QY 827 GYVKETDSNGVLTFFGANDIKGYETFDMSGFVAWVPVGASDNQDIRVAPSTAKKEGELT 886
Db 784 NLTKVADKDGFTFKGSEIKGYQVEVNGVLSVWVPVGAQSDQDIRVAASTRANAKGDKS 843
QY 887 LKATEAYDSQLIYEGSFNQFTIPDGSDDPSVYNNRKNIAENVDFKSGVGTSEFMAPQFVSA 946
Db 844 YTASQALDSQLIYEGSFNQDFV--QKDAQYTNKKIAENTDLFKAWGVTSEFMAPQYVSA 901
QY 947 DGTGFLDSYVONGYAFADRYDLAMSKNNKYGSKEDLRDALKALHAKAGIQAIADWVPDQIY 1006
Db 902 TDGTFJLDSIIQNGYAFSDRYDLAMSKNNKYGSKEDLANALKALHAAGIQAIADWVPDQIY 961
QY 1007 QLPKGVEVTTATRDGAGRKIADAIIDHSLYVANSKSSGKDYOAKYGGEGFLAEELKAKYPEM 1066
Db 962 QLPKGVEVTTASRVNRYGRVKIDOPWYNKLYLANTKSSGKDFQAKYGGEGFLAEELKQYPEM 1021
QY 1067 FKVNMIISTGKPIDDSVKLKWKAEBYFNGTNTVLERGYYVLSDEATCKYFTVTKEGNFIFL 1126
Db 1022 FTKAMISTGKPIDDSVKLKEWSAQYFNGTNTVLGRGTDYVLSDEGTCYFTVNEKEGFLEPA 1081
QY 1127 QLTGKEKVTIGSSDCKGTYFCTSGTQAKSAFVTENGNTYFPDARCHVMTNSEYSPNGK 1186
Db 1082 VLTGDKKATGFTYNDGKGMTYFTTAGSOAKSDFTVYAGNTYFPDYTGHMVT---GPNGI 1137
QY 1187 DV---YRFLPNGIWLNSAFYIDANGNTYVNSKQMYKGGVTFKPDVSETDKDGKESKVKVF 1244
Db 1138 NTKFYFFLNGVWLKDVAWEDDRGRSVYKGTGVMTKSGRNNWFANMDSKQO---LRF 1193
QY 1245 RYFTNEGVMAKGVTVIDGTFQYGEDGFOAKDKLVTFK-GKTYYPDAHTGNGIKOTWRNI 1303
Db 1194 RHPDNYGFMGSVGLVTHGNVQYYDEBGFQVKGDFVTDKAGQTRYFDKNTGNLVKQGFQNFQ 1253
QY 1304 NKGWYIFDANGVAATGAQVINGOKLYFN-EDGSQVKGGVVKNADGYTSYKKGFGBELVTN 1362
Db 1254 NGHWYYSDDQGLIAKGAQTIKGOKLYFDKTAQKVGDFVTDKDGNTFFYSGDGTGLAVS 1313
```

```
QY 1363 EFFTTDGNVYVYAGANGKTTVTGAQVINGQHLFYNA-DGSQVKGGVVKNADGTYSKYNAST 1421
Db 1314 TFFSTGNNAWFYADENGHVAKGEKTINGKLYFDTKTGQAKGRFVRDAKG-LRFYDADT 1372
QY 1422 GERLTNEFTT--GDNWYIYGANGKSVTGEVKI-GDDTYFFPAKDGKQVKGQTVSAGNR 1478
Db 1373 GALVTNSPLETKAGSNQWYMGADGYAVRGHQTIQSRHMYFDAETGQQAQKGIYVVTDANGER 1432
QY 1479 ISYVYGDGSKRAVSTWIEIQPGVYVYVFDKNGLA 1511
Db 1433 KYFYDANTGDRVNVQFVLVN-GSYFFFGYDGAA 1464

RESULT 3
Q56CX8 9STRE PRELIMINARY; PRT; 1506 AA.
AC Q56CX8;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Glucosyltransferase-T.
GN Name-gtfT;
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1310;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=B13N;
RA Kuwahara N.;
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY966490; AAX76986.1; -; Genomic_DNA.
KW Transferase.
SQ SEQUENCE 1506 AA; 167016 MW; 94B882EF2C17C451 CRC64;

Query Match 49.8%; Score 3949.5; DB 2; Length 1506;
Best Local Similarity 52.6%; Pred. No. 2.4e-169;
Matches 811; Conservative 211; Mismatches 405; Indels 115; Gaps 24;

QY 1 MENKIHKLHKYKKQWVTIIVASVALATVGLSVTTSSVSAD-----TQDKTVTQ 52
Db 1 MERKLHKLHKYKKQWVTIIVASAGLASIVAGSL-SQTVSADDLAKAQAAASQKAAAN 59
QY 53 SNSGTASLTVSPEATKEADLTNTKEADVLTPAKETNAVETATTTNTQATAEA----- 106
Db 60 QNEDEVASDAATASAKATSEKEVQSSDT---NSETNOVETKDAQSAKESADAVAKQAP 116
QY 107 -ATTATTADVA---VAAVPNKEAVVTTDAPATVTEKABEOPATVKAEVVNTVEVKAPEAA 161
Db 117 QAGPATTSQVASSSESSSVAPSKEA-----DKAAAGS 147
QY 162 LKDSVEEAALSILKNIKNIDGKYIYVNEGDGSHKENFAITVNGQLLYP-GKDGALTSSSTYS 220
Db 148 VSQNEEEAALSILANIKKIDGKYIYVNEGDGSHKENFAITVNGQLLYPDAKTGALSSTYS 207
QY 221 FPGPTTNIVDGFSINNRAVDSSEASGELIDGVLTDADSWYRPAIIKDGVTWQASTAEDPR 280
Db 208 FSQGLTPIVSDFSVNNKAFDSSEKSEFELVDGLTRESWYRPAKILENGKTVWDSKETDLR 267
QY 281 PLLMAWPNVDVTQVNYLNTMSKVFNLDAKYSSTDKQETLKVAADQI QIKIEQKIQAEKST 340
Db 268 PVLMSWPNKDTQVAYLVNTMSKALGCKEBFTTKTSQTALNTAAELIQAEIARVSKEQGT 327
QY 341 QMLRETIQAFVKTQPMNKETENYSKGGEDHLOGGALLYVNDSTRTPWANSYRRLNRTA 400
Db 328 KWLREMAAFVATQSRWNNKDSQYDKA---DHLOGGALLYTNKLTEWANSNWRLLNRTP 384
QY 401 TNQGTGIDKSLDEQSDPNHMGGFPLLANDVDLSNPVVOAQLNQIHYLMWGSIVMGD 460
Db 385 TPQDGKTHYSKADK-----YGYEFLANDVNSNPVVOAQLNQIHYLMWGEIVMGD 438
QY 461 KDANFDGIRVADVNDVADMLQLTYNYFREYYGVNKSSEANLAHISVLEAWSLNDNHND 520
```

Db	439	KNANFDGIRVADVNVNADTLQLTYTNYFNSVYGVNKSQAALAHISVLEAWSVNDNDYQ	498
Qy	521	KTDGAALAMENKORLALLFSLAKPIKERTPAVSPLYNNTFNTQDEKTDWINKGSKAY	580
Db	499	DTNGAALAMNDGLRFLSLLYLRPINERTPGMSTLIKSEYGLTRTK-----	545
Qy	581	NEDGTGVKQSTIGKYNEKYGDSGNVYFIRAHNNVQDIIAEIIKKEINPKSDGFTTDAE	640
Db	546	-----NDKYGDTQPSYVVFRAHDSVETVIAQIIKEIDTDTGFTFLDQ	591
Qy	641	MKQAFEYINKMLSSDKYTLNANIIPAAAYAVMLQNMETITRVYVYGDLYTDDGHMETKSPY	700
Db	592	LQAQFEIYINKDWNVSNKHYTHYNIIPAAAYAVMLSNMESVTRVYVYGDLYTDDGGYMAKSPY	651
Qy	701	YDTIVNLMKSRIKYVSGGQ--AQRSWLPPTDGMQNSDV-ELYRNEVTVSVRYKDIWMT	757
Db	652	YDAINTLLRARIRYAAGGQIMEHNSYKPSAAMKAAPDAGNVLGNSEVLVSFRFGQDVM	711
Qy	758	ANDTEGSKYSRTSGQVTVLANNPKNLNDQSAKLNVEMGKIHANOKYRALIVGTADGINKNF	817
Db	712	ADMTGGKLAKTSGMFTLLSNPELELDVNEEIKVNVGKIHAQOAYRPLLLTTDDGLQKY	771
Qy	818	TSADAIAAGYVKETDSNGVLTFGANDIKGYETFDMSGFVAVMVPVPGASDNQDIRVAPST	877
Db	772	LNDS-----TKLTKIADKDGFIIFKGEIKGYQVEVNGYLSVWVPVGAQADQDIRVAPST	828
Qy	878	EAKKEGELTKATEAVDSOLIIYEGFSNFQTIIPDGSDPSVYTRKRIABENVDLFKSGWVTFS	937
Db	829	AAKGEKAKTYTASQALLESQIIYEGFSNFQDFV--QKDSQYTNKKIAENTDLFKWGVTSF	886
Qy	938	EMAPQVSADGCTFLDSVLONGVAFADRYDLAMSNNKYGSKEDRLDALKALHKAGIOAI	997
Db	887	EMAPQVVSADGCTFLDSIIENGVAFTDRYDLAMSNNKYGSKEDLANALKALHAAGIOAI	946
Qy	998	ADWVPDQIYQLPGKEVVTATRTDAGRKIADAIIDHSLYVANSKSGKDYOQAKYGGEFLA	1057
Db	947	ADWVPDQIYQLPGKEVVTASRVNDYGRVKVQDPLVEKLYLANTKSGKDFQAKYGGEFLA	1006
Qy	1058	ELKAKYPENKPNMISTGKPIIDDSVKLQWKAEYFNGTNVLRGVGYVLSDBATGKYFTV	1117
Db	1007	ELQKKYPENFTTKMISTGKTIDPSVKLKEWSAKYFNGTNVLRGTYDILSDGTYKYFTV	1066
Qy	1118	TKEGNPIPLQCTGCKERVITGFSDDGKITVFGTSGTQAKSAFVTFNGNTYYPDARGHMT	1177
Db	1067	NEKGDPLPASLTGNKDAKTFYNDGKGIYIYTTAGNKARSFAFTEAGNTYIYDYTHMT	1126
Qy	1178	NSEYSPN--GKDYRFLPNGLMISNAFYIDANGNTYLYNSKGOMYKGYTKFDSVETDKD	1235
Db	1127	-----GNVINTKPYFLPNGLMKDAIKODEGRSVYVYKTYGMVYKGRDNEWFAMTDSK	1182
Qy	1236	GKESKVKPRYFTNEGVMAGVTVIDGFTQYFGEDGFQAKDLVTPK-GKTYYPDAHTGN	1294
Db	1183	GQ----MRFRHFDTRYGFMSTGLVTINQNVQYVDENGFOVKGBFVTDQGTQYFYDQSGN	1238
Qy	1295	GIKDWRNLNGWYVFDANGVAATGAOVINGOKLYEN-EDGSQVKGVVKNADGTYSKYK	1353
Db	1239	LVKQGLNKDGNWYLLDDQGLVAKGAQTIKGQKLYFDTKTGVOVKGDFVTDKGNFTFFYS	1298
Qy	1354	EGFGLVTFNEFTTDCGNWYVYAGANGKTVTGAOVINGOHLFYNA-DGSQVKGVVKNADG	1412
Db	1299	GUTGDLILQOFTSGNNAFYADENGHVAKGAKTIRGQKLYFDTKTGQAKGFFIRDDKG	1358
Qy	1413	TVSKYNASTGERLTNBEFTT--GDNWYVYIGANGKSVTGEVKI-GDPTYFFAKDGQVKG	1469
Db	1359	V-RYVDADTGTLVNAFLETKAGSNQWYMGADGVAVKGNTIKNQHMYFDAETGQOAKG	1417
Qy	1470	QTVSAGNRLSYVYVYDGSKRAVSTWIEIQVGVVYVFDKGLA	1511
Db	1418	IIVTDANGRKRYFDYTFGSRVNVNQFVLVN-GNMYFFGYDGA	1458

```
Db 521 SLAVLTK-----QPQRIDLNLISESVNKERAND-----T 551
Qy 597 KYGDASGNVYFIRAHNNVQDIIAIIKKKEINPKSDGFTTIDAEKQAFIYNKMLSSD 656
Db 552 AYGDITPTYSFVRAHDSHVOTVIKIVKEKIDTNSDGYTFTLDQLDAFKIYNEDMAKVN 611
Qy 657 KKYTLNIPRAAYVLMOMETITRVVYDLYTDDGHYWKTSPPYDIIVNLMKSRIKYVS 716
Db 612 KTYTHYNIIPAYALLSNMESVPRVYDLYTDDGQYMAKSPYDYDAITMLQRIAYVS 671
Qy 717 GGOAQRSLYPLTDGKMDSVDELVRTNEVTSVRYGKDIMTANDETSKYSRTSGQVTLV 776
Db 672 GGQSEEVH-----KVNQNIILSSVRYGQDLMSADDTQTGDLRSLSGLVTLV 718
Qy 777 ANPKMLDOSAKLNVEMGKIHANQYRALIVGTADGINKFTSDADAIAAGYVKETDSNG 836
Db 719 SNDPNLDLGDS-LTVNKGRAHANQAVRPLILGTGQVQSYLKDS--TNIVKYTDANG 774
Qy 837 VLTFGANDIKGYETFDMSGFVAVWVPVPGASDNQDIRVAPSTEAKKEGELTLKATEYDSQ 896
Db 775 NLFTFTADDIKGYSTVDMSGYLAVWVPVPGAKDQDVRVAADTNQKADGK-SLKTSAAALDSQ 833
Qy 897 LIYEGFSNFOTIPDGSPPSVYTNRKIAENVDLFKSMGVTSFEMAPQFVSADDTGTFELDSVI 956
Db 834 VIYEGFSNFQDF--ANNADYTHNKKIAENADFPKLGITSFEMAPQIVSATDGSFLDSII 891
Qy 957 QNGYAFADRYDLAMSNNKYGSKEDLRDALKALHAKAGIAIADWVPDQIYQLPGEVVTVA 1016
Db 892 QNGYAFSDRYDLAMSNNKYGSKDDLALANLALHANGIQAIADWVPDQIYQLPGEVVTVA 951
Qy 1017 TRTDGAGRKTDALIDHSLVANSKSGKDYQAKYGGEFPLAELKAKYPENFKNMIMSTGK 1076
Db 952 KRTNSYGNPTFDIAYINNALYATNTKSSGSDYQYGGAFDELKAKYPDMFTVMIMSTGK 1011
Qy 1077 PIDDSVKLKQWKAIEYFNGTNVLRGCVYLSLSDATGKYFTVTKEGNFIPLQLTKGKVIIT 1136
Db 1012 PIDPSTKIKQWEAKYFNGTNVLRGAGYVLSLSDATGKYFTVNSGDFPLPASFTGDDQNAKT 1071
Qy 1137 GFSSDGKGIYFTGTSCTQAQSAFVTFNGNTYTFDARGHVMVTNSEYSPNGKDVYRFLPNGI 1196
Db 1072 GFYDGTGMAYYSTGNKAVNSFIYEGGHYFYFDKGHWYTGYSYKAEDGNDYY-FLPNGI 1130
Qy 1197 MLSNAPYIDANGTYLINSKGQMYKGG--YTKFVDSVETDKGKESKVVKFRYPTNBSGVA 1254
Db 1131 QMRDAIYQDAQGNSYTYGRTGILYKGNWYFPFVDPNNANK-----TVFRYFDANNVMA 1183
Qy 1255 KGVTVIDGFTQYFEGDGFQAKDKLVTFKGTYYFDAHTGNGIKDWTNRNKGWYFEDANG 1314
Db 1184 IGRNNYGTQYYFDENGFQAKGQLLTDGKTHYFDEBNGAMAKNFVNVGDDWYMDNG 1243
Qy 1315 VAATGAQVINGQKLYFN-EDGSQVKGGVKNADGTYSKYKEGFGELVTNEFFTTDGNVWY 1373
Db 1244 NAVKGQYPVNNQILYFNPETGVQVKQFITDAQGRTSYDANSGLAKSSGFFTFPNSGDWY 1303
Qy 1374 YA-----GANGKTVTGAQVINGQHLYFNAD-GSQVKGGVKN 1409
Db 1304 YAENGYYVYKGFQVAENQDQWYFDTGTGQAKGAARVGDRLYFNPDSGVQVKGFATD 1363
Qy 1410 ADGTYSKYNASTGERLTNEFFTTGDNWYIYGANGKSVTGEVKI----- 1453
Db 1364 ESGNTSFYHGDNDKVVGGFFFTTGNNAWYADNNGNLVKGFBQIDGKWHYFDEVTGQOAK 1423
Qy 1454 -----GDDTYFPFADKGQVKGQVTSAGNRISYYGDSGKRAVSTWIEIQGVVYVFDK 1507
Db 1424 GAALVNGQQLYFDVDSGIQVKGDFVTDGQNTSYDVSNGDKKNGFFTTGDNWYIADG 1483
Qy 1508 NG 1509
Db 1484 OG 1485
```

RESULT 5
Q00599_STRSL

```
ID Q00599_STRSL PRELIMINARY; PRT; 1599 AA.
AC Q00599;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-FEB-2005 (T-EMBLrel. 29, Last annotation update)
DE Glucosyltransferase S.
GN Name=gtfk;
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 25975;
RX MEDLINE=93381463; PubMed=8371114;
RA Giffard P.M., Allen D.M., Milward C.P., Simpson C.L., Jacques N.A.;
RT "Sequence of the gtfK gene of Streptococcus salivarius ATCC 25975 and
RT evolution of the gtf genes of oral streptococci.";
RL J. Gen. Microbiol. 139:1511-1522(1993).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 25975;
RX MEDLINE=92148377; PubMed=1838391;
RA Giffard P.M., Simpson C.L., Milward C.P., Jacques N.A.;
RT "Molecular characterization of a cluster of at least tow
RT glucosyltransferase genes in Streptococcus salivarius ATCC 25975.";
RL J. Gen. Microbiol. 137:2577-2593(1991).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 25975;
RA Jacques N.;
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
RN [4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 25975;
RA Giffard P.M., O'Connor D.M., Milward C.P., Simpson C.L., Jacques N.A.;
RT "Sequence of the gtfK gene of Streptococcus salivarius ATCC 25975 and
RT the evolution of the gtf genes of oral streptococci.";
RL Mol. Microbiol. 0:0-0(1992).
DR EMBL; Z11872; CAA77898.1; -; Genomic DNA.
DR EMBL; Z11873; CAA77901.1; -; Genomic DNA.
DR PIR; S22737; S22737.
DR HSSP; P06653; 1H8G.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 3.
DR Pfam; PF02324; Glyco_hydro_70; 1.
DR Transference.
SQ SEQUENCE 1599 AA; 176480 MW; 24B77869E152B707 CRC64;

Query Match 47.2%; Score 3743; DB 2; Length 1599;
Best Local Similarity 47.1%; Pred. No. 5.3e-160;
Matches 787; Conservative 234; Mismatches 417; Indels 232; Gaps 26;

Qy 1 MENKHYKLHKVKQWVTIIVASVALATVGLGLSVTTSSVSADETODKTVTQNSGTTAS 60
Db 1 MENKRYKLHKVKQWVTIIVASVALATVGLGLSVTTSSVSADETODKTVTQNSGTTAS 58
Qy 61 LVTSPEATKADKRTNWKADVLTPAKETNAVETATTTNTQATAEAAATTA-----TTADV 115
Db 59 ENTNPVVEKEVGTT-----EVANTSNNATTTTERRAEVTDKPKAGTTVQP 101
Qy 116 AVAAPNKEAVTTDAPAVTTEKAE-----EQATVKAEEVNVTEVKAP-----EALAKDSEV 167
Db 102 NSGTTSDRAAAVEVEAKPETTAKPEVATKPEATTATTSSEVAANAGVAAPTTEKSKELSEAEI 161
Qy 168 EAAALSLKNI-KNIDGKYVYVNDGSHKENFALTIVNGQLLYFGKGALTSSTSYSPFTGTT 226
Db 162 KAAVSLDNKKKGGKYYLLLEDGSHKGNFALTIVNGQVLYFDENGALSTSTSYFTQETT 221
Qy 227 NIVDGF SINNRAYDSSEASFELIDGLTADSWYRPASIIKDGVTWQASTAEDPRPLIMAW 286
```



```
QY 601 ASGNYVPIRAHDNNVDIIAIIKEINPKSDGFTTDAENKQAFETYNKDMLSKKYIT 660
Db 618 RTANYIFVRADHSEVQTVIADIIRENINPTDGLTFTMDLQKAFKYNEDMRKADKYIT 677
QY 661 LNNIPAAVAMQNMETITRVYIGDLVTDGHHMETKSPYYDTIVNLKMSRIKYVSGQA 720
Db 678 QFNIPTAHALMLNKSITRVYIGDLVTDGQWMEKSPYHDAIDALLRARIKYVAGGD 737
QY 721 QRSYVPTDGMKNSDVVELYRTNEVTSVRYGKDIIMTANDTEGSKYSRTSGQVTLVANNP 780
Db 738 MKVYMGVPREADK----WSYNGILTSVRYGTGANEATD-EGTAETRTQGMVIAASNP 791
QY 781 KLNLDOSAKLVNEMGKIHAKQKTRALIVGTADGINKFTSDADAIAAGYVKETDSNGLVTF 840
Db 792 NLKLNMDKLVQVNGAAHKQYRPVLLTTKDGISRYLTD-BEVPOSMLKKTDTANGILTF 850
QY 841 GANDIKGYETFDMSGFVAVWVPVPGASDNQDIRVAPSTEAKKEGELTLKATEAVDSOLIVE 900
Db 851 DMMDIAGYVNVQSGYLAWVPVGAKEHQDARVTASKKQNASQV-YESSPALDSOLIVE 909
QY 901 GFSNFQTIPIGSDPSVYTNRKIAENVDLFKSMGVTSEFEMAPQFVSADGDTFLDSVIONGY 960
Db 910 GFSNFQDFATRDQ--YTNKVIKVNLPKEMGVTSFELPPQVVSQDGTFLDSIIIONGY 967
QY 961 AFADRYDLAMSKNKYSGKEDLRDALKALHAKIGIAIADWVPQIYQLPGKEVVTATRTD 1020
Db 968 AFEDRYDMAKSKNKYSGSLDRLNALRALHSHVNIQAIADWVPQIYNLPKGEVVTATRVN 1027
QY 1021 GAGRKIADAIIDHSLVANSKSGDYQAKYGGSEFLAELKAKYPENPKVNMISTGKPIDD 1080
Db 1028 NYGTYREGAEIEKELVANSKKTNGTDYQGYGAFIDELKAKYPEIFERVQISNGQKMTT 1087
QY 1081 SVKLQWKAIEYFNGTINVLBERGVYVLSDEATGKYFTVTKEGNFIPQLTGKEKVIITGFS 1140
Db 1088 DEKITKSAKYFNGTILGRGAYVVLKDWGSKYELSNKNGETALPKQLVNKE-ASTGFVK 1146
QY 1141 DKGITYFTGSGTQAKSAFV-TPNGNTYFPDAGHMTNSEYSPNGKQVYRFLPNGIMLS 1199
Db 1147 DTNGFKFYSTSGNQAKDTFTQDENGWNYFDNQGLVLTGAR-EIDGKQLY-FMKGNGVQLR 1204
QY 1200 NAFYIDANGNTVLYNSKGQYKGYTFKDFVSETDKDGESKVVKPRYFTNEGVMAGVTV 1259
Db 1205 DALQDENGQYDYDKTGAKVLNRY-----YTSQG-----NMRYPDAGVMARGLVK 1252
QY 1260 IDGFTYFGEDGFQAKDKLVTFK-GKTYFYFDAHTGNGIKDTRNIN--GKWTYFDDANGVA 1316
Db 1253 IDGGQYFPDQGVQVGVKVRADGKLRVFDKDSGNVLRNRFAGQGNPDSDWYFPGADGVA 1312
QY 1317 ATGAQVINGOKLYFNBEDGSQVGVVKNADGYSKYKEGFGELVTNEFFTTDGNVWYAG 1376
Db 1313 LTGLQKIGQOTLYFGQDGKQVGVVMLADKSIYFDANSGEMAVNKAFAEGAKNEWYFD 1372
QY 1377 ANGKVTGQAVINGQHLVFNADGSQVGVVKNADGYSKYNASGTERLTNEFFTTGDN 1436
Db 1373 QDQKAVTGKLTINNQLYFPDQGVQVGVVLTADKSIYFDANSGEMAVNKAFAEGAKNE 1432
QY 1437 WYVIGANGSVTGEVKIGDDTVFFAKDGQVKGQVTSAGNGRISYYVYDGSGRVASTWIE 1496
Db 1433 WYTFDQDGAHVTLQTLNQVLYFGQDGKQVGVVLTADKSIYFDANSGEMAVNKAFAE 1492
QY 1497 IQPGVYVYFPKNGLA 1511
Db 1493 GAKNEWYFDDGKA 1507
```

RESULT 8

```
Q9LCH3 STOR PRELIMINARY; PRT; 1575 AA.
AC Q9LCH3.
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Glucosyltransferase.
```

```
GN Name=gtrf;
OS Streptococcus oralis.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OC NCBI_TaxID=1303;
RN [1];
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC10557;
RX DOI=10.1128/IAI.68.5.2475-2483.2000;
RA Fujiwara T., Hoshino T., Sobue S., Hamada S.;
RT "Purification, characterization, and molecular analysis of the gene
encoding glucosyltransferase from Streptococcus oralis.";
RL Infect. Immun. 68:2475-2483(2000).
DR EMBL; AB025228; BAA95201.1; -; Genomic_DNA.
DR HSSP; P06653; 1HCX.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW:binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW:binding_1; 4.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1575 AA; 176792 MW; 772A26E4D7C2E543 CRC64;
```

```
Query Match 42.3%; Score 3350.5; DB 2; Length 1575;
Best Local Similarity 44.2%; Pred. No. 2.6e-142;
Matches 723; Conservative 227; Mismatches 491; Indels 193; Gaps 30;
```

```
QY 1 MENKIHKLHKVKQWVTIIVASVALATVGLGLSVTSSVSADETQKTV---TQNSG 56
Db 2 MEKKIHKMKHKQKQWVALATLTIIVAPKVLGLSGVHVHADDVKQVVQBEPATQTS 61
QY 57 -----TTASLVTSPEATK--EADKRTNTKEADVLTPAKETNAVETATTNTQATAEAAT 108
Db 62 PQQTPAQAKIASQEAEEKVTPADKVTDDVAASE-KPAKPAENTEATVQTNAQEPKPAD 120
QY 109 T--ATTADVAVA-----AVPNKEAVVTTDAPAVTTEKAEQOPATVKAEVNVTEVK 156
Db 121 TREASTERKAAVAEEVKANNAITEIPKTEVADQNKQARPTAQDQEDGKEKTAVEKIVA 180
QY 157 APEAALKD-----SEVEAALS-----LKNKNIDGKYYYVNE 190
Db 181 NPKVAKKRLPFGSKQGAIAERMVADQAPVNAVHADHDDVLSHIKTI 240
QY 191 SHKENPAITVNGQLLYF-GKQALTSSTYSFTPGTNTIVDGFSGINNRAVDSEASFELI 249
Db 241 TVKKNFAVELNRIILYFAETGALVDSNEYQFQOGTSSLNNEFBSQKNAFYGTDKDIETV 300
QY 250 DGYLTADSWYRPAASIIKQGVWQASTAEDFRLLMAWPNVDTQVNLNYSKVFNLD 309
Db 301 DGYLTADSWYRPFILKQKVTASTETDLRFLMAWPDKRTQINLYNMQOGLGAGA 360
QY 310 YSSTDQKQTLKVAADKIQIKI EOKIQAESTQWLRETSFAFKTQPMNKETENTYSGGG 369
Db 361 PENKVEQALLTGASQVQVKIEKIGEGDTKWLRTLMGAFVKTQPNWNIKTESETTGK 420
QY 370 EDHLQGGALLYNDERTWANSDYRLNRTATNQGT-----IDKSILDSQSDPNHMGFPD 425
Db 421 KDHLQGGALLYTNNEKSPHADSFKFLNRTPTTSQTGTPKYFDKS-----NGGYE 470
QY 426 FLLANDVLSNPPVQAEQLNOIHYLMNWGSIYVWGDKDANFDGIRVDADVNDADMLQLVT 485
Db 471 FLLANDFNSNPVQAEQLNHLHYMNFSGIVANDPTANFDGVRVDADVNVNADLLQIAS 530
QY 486 NYFREYGVNKSSEANALAHISVLEAWSLNDHNDKTDGAALAMENKORLALLFSLAKPI 545
Db 531 DYFKSRKYVGESEBEAIIKHLISLEAWSNDPDYNDKDTGAQLAIDNKLRLSLYSFMRNL 590
QY 546 KERTPAVSPLYNNTTQREKTDWINKDGSKAYNEDGTVKQSTIGKYNKYEKQASGNY 605
Db 591 SIRS-GVEPTITNSLNDRSSSEKKN-----GERMANY 620
```

QY 606 VFIRAHNNVODIIIEIIEIKKEINPKSDGFTITDAEMKQAFELYNKXMDLSSDKKYLNNIP 665
Db 621 IFVRAHSEVQVIADIIIEININPNDGLFTTMDLQKAFKLYNEDMRKADKKYTFQFNP 680
QY 666 AAYAVMLQNMETITRYVYGDLYTDGCHYMETKSPYDITVNLMSRIKRYVSGQQRYSW 725
Db 681 TAAHALMSNKSDITRYVYGDLYTDGQYMEKSPYHDAIDALLARIKIVVAGQDMKVY 740
QY 726 LPTDGHKXNSDELVTNEVTSYRYGKIDIMTANDTEGSKYSRTSGQVTLVANNPKLND 785
Db 741 MGVPREADK----WSYNGILTSYRYGTGANEATD--EGTAETRTQGMVIANNNPKLN 794
QY 786 QSAKLNVGKIHANOKYFALVGTADGJKNFTSDADAIAAGVVKETDSNGVLTGANDI 845
Db 795 EWDKUGVNGAHHKQYRPVLLTTKDGISRYLTD--EEVPSQLWKKTDANGILTFDMNDI 853
QY 846 KGYETFDMSGFVAVVPVUGASNDQDIRVAPSTEAKKEGELTLKATEAYDSQLIYGFNSF 905
Db 854 AGSYNVQVSGYLAVVPVUGAKADQDARTTASKKNASQGV--YESSAALDSQLIYGFNSF 912
QY 906 QITPDGSDPSVYTRNKIAENVDLFKSGVTSFEMAPQFVSADDTGLDVSIVONGVAFADR 965
Db 913 QDPATRDQ--YTNKVIKKNVLFKEWGVTSFELPPQYVSSQDGTFLDSIIQNGYAFED 970
QY 966 YDLAMSKNNKYSKEDRLDALXALHKAGIQATADVDPDQIYOLPGKEVVTATRTDAGRK 1025
Db 971 YDMAMSKNNKYSGLDLLALXALHSVNIQALADVDPDQIYNLPKGVEVVTATRVNNGY 1030
QY 1026 IADAIIDHSLYVANSKSGDKYQAKYGGFELAEKAKYPEMPKVMNISGPKPIDDSVKLK 1085
Db 1031 REGAEIKEKLYVANSKTNEDTQKYGGAFLDELAKYPEIFERVQISNGQKMTTDEKIT 1090
QY 1086 QWKAEEYFNGTNVLERGVYVLSDEATGKYFTVTKEGNFIPLOLTGKEKIVITGFSSDGKI 1145
Db 1091 KWSAKYFNGTNILGRGAYVYLKDWASNDYLTNRNGEIVLPKQLVNKS--YTFGVSDANGT 1149
QY 1146 TYFGTSGTQAKAFV--TFNGNTYFDPARGHMTNSEYSPNGKDVYFELNGMLSNFVY 1204
Db 1150 KYFSTSGYQAKNSFIODENGNNWYFDPKRGYLVGTGA--HEIDGKHVY--FLKNGIQLRDSIRE 1207
QY 1205 DANGNTLYNSKG--QMYKGYTKFDVSETDKDKESKVKVFRYFTNEGVMAGKVTVIDGF 1263
Db 1208 DENGNYYYDQGAQVLRNYT-----TDGQ-----NRYFDAQVMARGLVKIGD 1254
QY 1264 TQVFGDGFQAKDLVTFK--GKTYFPAHTGNGI-----KOTW----- 1300
Db 1255 QQFFDENGQVKGKIYSAKDGKLRYPDKDSGNVNRPAQGNPDSWYFVGVEFAKLTLGL 1314
QY 1301 -----RNNGK----- 1315
Db 1315 QKIGQOTLYPDQDGKQVKGKIVTLSDKSTRYFDANSSEMAGKFAEGAKNEWYFDKTK 1374
QY 1316 AATGAQVINGQKLYFNEEDSGQVKGKYNADGTYSKYKEGFLVNEFTFTDGNWVYA 1375
Db 1375 AVTGLQKIGKQTLYFDQDGKQVKGKIVTLADKSIYRFDADSGEMAGKFAEGAKNEWYF 1434
QY 1376 GANGKTVTGAQVINGQKLYFNEADSGQVKGKYNADGTYSKYNASTGERLTNEFTFTGN 1435
Db 1435 DQTKRAVTGLQKIDKQTLYFDQDGKQVKGKIVTLSDKSTRYFDANSSEMAGKFAEGAKNEWYF 1494
QY 1436 NMYIYGANGKSTVGEVKIGDITYFFAKDGKQVKGKQVTSAGNRI SYYYGDSGKRAVSTWI 1495
Db 1495 EWYFDPQAKAVTGLQVQOQTLYFTQDGKQVKGKVDV--NGVSRVFDANSMDMAKSWI 1553
QY 1496 ETQPGVYVYFDKNG 1509
Db 1554 QLEDGSWYVFDKNG 1567

RESULT 9
Q59983 9STRE
ID Q59983 9STRE PRELIMINARY;
AC Q59983 9STRE PRT; 1590 AA.

DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-WAR-2004 (Tremblrel. 26, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5).
OS Name=gtf1;
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_taxid=1310;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC MEDLINE=94146405; PubMed=8312602;
RA Sato S., Inoue M., Handa N., Aizawa Y., Isobe Y., Katayama T.;
RT "DNA sequence of the glucosyltransferase gene of serotype d
Streptococcus sobrinus";
RL DNA Seq. 4:19-27 (1993).
RN [2]
RP PROTEIN SEQUENCE.
RX MEDLINE=91224988; PubMed=1827439;
RA Mooser G., Hetta S.A., Paxton R.J., Shively J.E., Lee T.D.;
RT "Isolation and sequence of an active-site peptide containing a
catalytic aspartic acid from two Streptococcus sobrinus alpha-
glucosyltransferases";
RL J. Biol. Chem. 266:8916-8922 (1991).
DR EMBL; D13858; BAA02976.1; -; Genomic_DNA.
DR PIR; A39841; A39841.
DR HSP; P06653; 1HCX.
DR GO; GO:0047849; F:dehydratase activity; IEA.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. .; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding_70.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Glycosyltransferase; Signal; Transferase.
FT SIGNAL 1 38 Potential.
FT CHAIN 39 1590 glucosyltransferase-I.
SQ SEQUENCE 1590 AA; 175956 MW; C3C83A57CF3C2B0B CRC64;
Query Match 41.9%; Score 3325.5; DB 2; Length 1590;
Best Local Similarity 45.0%; Pred. No. 3.5e-141;
Matches 710; Conservative 239; Mismatches 456; Indels 173; Gaps 33;
QY 1 MENKHYKLHKVKKQWTVIAVASVA-LATVIGLGLSVTTSSVSADTQKTVTQNSGTTA 59
Db 1 MEKNVRFKMKHYKKEVWTLVSASATMLASALG-----ASVASAD-----TDTASDD---- 46
QY 60 SLVTSPEATKEADKETNTKEADVLTPAKETNAVETATTTNTQATAEAATATTADVAVA 119
Db 47 ----SNQAVVTGDTNNQATD-----QTSIAATAT-----SEQSASTDAATQASAA 90
QY 120 VPKNEAVVTTDAPAVTTEKAEQPAVTKAEVNV-----TEVKAPAAALKDSVEEAL 171
Db 91 EQTQGTASTDTAAQTNTTANB-AKWVPTENENQGTDEMLAEAKNVATAESDSIPSDLA 149
QY 172 SLKNTKNDIGKYYVNEGSHKENPAITVNGQLLYFGKDGALTSSTTSYVFTGTTNIVDG 231
Db 150 KMSNVKQVGDQYKYVYDQDGNVKNFAVSGDKIYVYFDETGAYKDTSKVDADSSAVSQN 209
QY 232 ---FSNNRNVDSSEASFEIDGYLTADSVWRPASIKDGVTWQASTADPRPLMAWMP 288
Db 210 ATIFANNRNVYSTSAKNFEAVDNYLTADSVWRPKSILKDGKTWTSGDKDPRPLMAWMP 269
QY 289 NVDTQVNYLNMYSKVFNLDAKYSSTDQKTELKVAARDIKIEQKIQAEKSTQWLRETS 348
Db 270 DTEKRNVYVNMKVVGIDKTYTAETSQADLTAAAEVQAEQKISNTNNKWLRETS 329
QY 349 AFVKTQPOWNETEN-YKGGGEDHLQGGALLVYVDS-RTPWANSYRRLNRATNTQGT 406
Db 330 AFVKTQPOWNGESEKPY-----DDHLQNGALLFDNQTLTPTQSYRLLNRPTNQTS 384
QY 407 IDKSLDEQSDPNHMGGFDFLLANDVDLSNPVVAEQLNQIHYLMNMGSI VMGDKDANFD 466

874 TDGQFLDSVITQNGYAFTDRYDLGMSKANKYGTADQLVKAIKALHAKGLKVMADWVPDQMY 933

1007 QLPKEKEVWTRTDGAGRIADIADHSILYVANSKSGKDYQAKYGGEFIAELKAKYPEM 1066

934 TFPQKEVWTVTRDKFGKPIAGSINHSLVYVTDTKSGDDYQAKYGGAFDDELKEKYPEL 993

1067 FKNVMIISTGKPIIDDSVKLQWKAEYFNGTNVLERGVGYVLSDBATQKIFVTTKBGNFIPL 1126

994 FTKKQISTGQAIDPSVKIKQWSAKYFNGSNILGRGADYVLSDDQVSNKYFNVASDITLFLPS 1053

1127 QLTGKEKVIITGFSDDGKITY-FGTSGTQAKSAFVTFNGNTYVYFDARGHWNVTSEYSPNG 1185

1054 SLUGK-WBESGIRYDGKGIYINSSATGDQVKASFITEAGNLVYFGKDGVMYVTGAQ-TING 1111

1186 KDYRFLPGLMISLNFAPYIDANGNTLYLNSKGMQYKGGVTKFDVSETDKDGKESKVVKFR 1245

1112 AN-YFLENGTALRNTIYTDAGNSHYANDGKRYENGQQP-----GND-----WR 1157

1246 YFTNEGMAKGVTVIDGFTQYFGBEDGFAQOK-LVTFKGTYYVYFDAHTGNGIKDWTW-RMI 1303

1158 YF-KDGNMAVGLTVDGNTQYFDFKGVQAKDKIIVTRDGKRVYFDQHNNGNAVNTFIADK 1216

1304 NGKWWYFDANGVAATCAOVINGOKLYFNEBDSGVKGVVKNADGTYSKYKEGFELVTNE 1363

1217 TGHVYLGKDGVAVTGAQTVGKQKLYFEANGBQGVDFVTSHEGKLYFYDVSDDGMWTD 1276

1364 FFTTGDGNVYAGANGKTVTGAOVINGOHLNFNADGSQVKGVVKNADGTYSKYNASTGEB 1423

1277 FIEDKAGNFYLGKDGAAVSGAQITRGKLYPKAYGQQVKGDIIVKGTDKIRYYDAKSGE 1336

1424 RLNE-----FFTTG 1433

1337 QVENKTKYKAADGKTYVIGNNGVAVDPSVVKGTQFKDASGALRFYNLKGQLVGTSGWYETA 1396

1434 DNNWYIYGANGKSVTGEVKGIDDTYFPKADGQVKGQVTSAGNGLRISYYIGDSGKRAVST 1493

1397 NHDWVYI-OSGKALTEGEQTNGOHLFYKEDGHQVKGQLVTRTDGKRVYRYDANSGDQAFNK 1455

1494 WIEIQPGVVVYFDKNGLA 1511

1456 SVTVNGKTY-YFGNDGTA 1472

RESULT 12

GF2_STRDO

ID_GTF2_STRDO STANDARD; PRT; 1592 AA.

AC P27470;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 13-SEP-2005 (Rel. 48, Last annotation update)

DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)

DE (Sucrose 6-glucosyltransferase).

OS Streptococcus downei (Streptococcus sobrinus).

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

NCBI_TaxID=1317;

[1]

RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].

RC STRAIN=6715 / Serotype G;

RX MEDLINES=91123227; PubMed=1704006;

RA Abo H., Matsumura T., Kodama T., Ohta H., Fukui K., Kato K., Kagawa H.;

RA "Peptide sequences for sucrose splitting and glucan binding within

RT Streptococcus sobrinus glucosyltransferase (water-insoluble glucan

RT synthetase).";

RL J. Bacteriol. 173:989-996(1991).

-/- FUNCTION: Production of extracellular glucans, that are thought to

CC play a key role in the development of the dental plaque because of

CC their ability to adhere to smooth surfaces and mediate the

CC aggregation of bacterial cells and food debris.

CC -/- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-

CC fructose + (1,6-alpha-D-glucosyl) (n+1).

CC -/- SUBCELLULAR LOCATION: Secreted.


```

RESULT 13
GTFB STRMU
ID GTFB STRMU STANDARD; PRT; 1476 AA.
AC P08987; O69381; O69384; O69390; O69396;
DT 01-NOV-1988 (Rel. 09, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
GN Name=gtfB; OrderedLocusNames=SMU.1004;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OC NCBI_TaxID=1309;
OX [1]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=GS-5;
RC MEDLINE=87308013; PubMed=3040685;
RX Shiroza T., Ueda S., Kuramitsu H.K.;
RA "Sequence analysis of the gtfB gene from Streptococcus mutans.";
RL J. Bacteriol. 169:4263-4270(1987).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MT4239 / Serotype c, MT4245 / Serotype e, MT4251 / Serotype f,
RC MT4467 / Serotype e, and MT8148 / Serotype c;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
RN [3]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=UA159 / ATCC 700610 / Serotype c;
RX MEDLINE=22255063; PubMed=12397186; DOI=10.1073/pnas.172501299;
RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,
RA Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
RT pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
CC -I- FUNCTION: Production of extracellular glucans, that are thought to
CC play a key role in the development of the dental plaque because of
CC their ability to adhere to smooth surfaces and mediate the
CC aggregation of bacterial cells and food debris.
CC -I- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-
CC fructose + (1,6-alpha-D-glucosyl) (n+1).
CC -I- SUBCELLULAR LOCATION: Secreted.
CC -I- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes
CC water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
CC forms of glucans.
CC -I- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
CC -I- SIMILARITY: Contains 10 cell wall binding repeats.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL; M17361; AAA88588.1; -; Genomic DNA.
CC EMBL; D88651; BAA26101.1; -; Genomic DNA.
CC EMBL; D88654; BAA26105.1; -; Genomic DNA.
CC EMBL; D88657; BAA26109.1; -; Genomic DNA.
CC EMBL; D88660; BAA26113.1; -; Genomic DNA.
CC EMBL; D89977; BAA26119.1; -; Genomic DNA.
CC EMBL; AS014940; AAN58705.1; -; Genomic DNA.
CC FIR; B33135; B33135.
CC HSSP; P06653; IH8G.
CC InterPro; IPR002479; Cell_wall_bd_put.

```

```

DR InterPro; IPR003318; Glyco_hydro_70.
DR PFam; PF01473; CW_binding_1; 4.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Complete proteome; Dental Caries; Glycosyltransferase; Repeat; Signal;
KW Transferrase.
FT SIGNAL 1 34 Potential.
FT CHAIN 35 1476 Glucosyltransferase-I.
FT REPEAT 1097 1130 A repeat.
FT REPEAT 1161 1210 1.
FT REPEAT 1225 1275 2.
FT REPEAT 1290 1340 3.
FT REPEAT 1355 1405 4.
FT REPEAT 1420 1470 5.
FT REGION 35 1051 Catalytic (approximate).
FT REGION 1097 1476 Glucan-binding (approximate).
FT REGION 1161 1470 5 X tandem repeats.
FT VARIANT 62 62 S -> T (in strain MT4239).
FT VARIANT 65 65 T -> I (in strain GS-5).
FT VARIANT 68 68 V -> A (in strain GS-5, strain MT4245, strain MT4251, strain MT4467 and strain MT8148).
FT VARIANT 78 78 Q -> P (in strain MT4251).
FT VARIANT 86 86 I -> S (in strain GS-5, strain MT4245, strain MT4251, strain MT4467 and strain MT8148).
FT VARIANT 89 89 S -> F (in strain MT4251).
FT VARIANT 168 168 K -> N (in strain MT4251).
FT VARIANT 276 276 S -> D (in strain GS-5, strain MT4467 and strain MT8148).
FT VARIANT 399 399 N -> R (in strain MT4239).
FT VARIANT 474 474 I -> T (in strain MT4239).
FT VARIANT 512 512 K -> R (in strain MT8148).
FT VARIANT 519 519 F -> Y (in strain MT8148).
FT VARIANT 701 701 T -> I (in strain MT8148).
FT VARIANT 708 708 A -> V (in strain MT8148).
FT VARIANT 938 938 F -> L (in strain MT8148).
FT VARIANT 952 957 FGPVE -> YGTPVA (in strain GS-5, strain MT4239 and strain MT4467).
FT VARIANT 963 964 SV -> NT (in strain GS-5, strain MT4239 and strain MT4467).
FT VARIANT 968 970 ADS -> VDG (in strain GS-5, strain MT4239 and strain MT4467).
FT VARIANT 1086 1086 A -> T (in strain MT4239).
FT VARIANT 1158 1158 S -> N (in strain MT4239).
FT VARIANT 1163 1163 H -> Y (in strain MT4251).
FT VARIANT 1168 1168 E -> K (in strain MT8148).
FT VARIANT 1182 1182 Y -> C (in strain MT8148).
FT VARIANT 1234 1234 A -> P (in strain MT4239).
FT VARIANT 1263 1263 R -> H (in strain GS-5 and strain MT4467).
FT VARIANT 1263 1263 R -> P (in strain MT8148).
FT VARIANT 1264 1264 Y -> H (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).
FT VARIANT 1272 1272 S -> G (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).
FT VARIANT 1329 1329 H -> Y (in strain GS-5 and strain MT4467).
FT VARIANT 1394 1394 Y -> H (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).
FT VARIANT 1402 1402 S -> G (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).
FT VARIANT 1459 1459 Y -> H (in strain MT4467).
FT VARIANT 570 570 R -> A (in Ref. 1).
FT CONFLICT 800 817 ADQDVRVAASTAPSTDGK -> LIRKFAURLARPHQOMA (in Ref. 1).
FT CONFLICT 1310 1310 H -> L (in Ref. 1).
FT CONFLICT 1476 1476 AA; 165847 MW; 9C6E09F731B4CBFC CRC64;
SQ SEQUENCE

```

```

Query Match 40.9%; Score 3239; DB 1; Length 1476;
Best Local Similarity 45.5%; Pred. No. 2.5e-137;
Matches 700; Conservative 228; Mismatches 474; Indels 136; Gaps 30;
QY 1 MENKHYKLHKVKQWTVIAVAS--VALATVLGGLSVTTSSVSADFTQDKTVTSNGTT 58

```


[4] NUCLEOTIDE SEQUENCE OF 1-349.
 RN STRAIN=GS-5;
 RC MEDLINE=87308013; PubMed=3040685;
 RA Shiroza T., Ueda S., Kuramitsu H.K.;
 RT "Sequence analysis of the gtfB gene from *Streptococcus mutans*.";
 RL J. Bacteriol. 169:4263-4270(1987).
 CC -|- FUNCTION: Production of extracellular glucans, that are thought to
 CC play a key role in the development of the dental plaque because of
 CC their ability to adhere to smooth surfaces and mediate the
 CC aggregation of bacterial cells and food debris.
 CC -|- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-
 CC fructose + (1,6-alpha-D-glucosyl) (n+1).
 CC -|- SUBCELLULAR LOCATION: Secreted.
 CC -|- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
 CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes
 CC water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
 CC forms of glucans.
 CC -|- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
 CC -|- SIMILARITY: Contains 5 cell wall binding repeats.
 CC -----
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use as long as its content is in no way modified and this statement is not
 CC removed.
 CC -----
 CC EMBL; M22054; AAA88592.1; -; Genomic DNA.
 CC EMBL; D88652; BAA26102.1; -; Genomic DNA.
 CC EMBL; D88655; BAA26106.1; -; Genomic DNA.
 CC EMBL; D88658; BAA26110.1; -; Genomic DNA.
 CC EMBL; D88661; BAA26114.1; -; Genomic DNA.
 CC EMBL; D89978; BAA26120.1; -; Genomic DNA.
 CC EMBL; A8014940; AAN58706.1; -; Genomic DNA.
 CC EMBL; M17361; AAA88589.1; -; Genomic DNA.
 CC FIR; JT0343; JT0345.
 CC HSSP; P06653; IH8G.
 CC InterPro; IPR002479; Cell wall bd put.
 CC Pfam; PF01473; CW_binding_1; 2.
 CC Pfam; PF02324; Glyco_hydro_70; 1.
 CC Complete proteome; Dental caries; Glycosyltransferase; Repeat; Signal;
 CC Transferrase.
 CC SIGNAL 1 34
 CC CHAIN 35 1455 Glucosyltransferase-SI.
 CC REPEAT 1126 1159 A repeat.
 CC REPEAT 1169 1200 A repeat.
 CC REPEAT 1227 1238 C repeat.
 CC REPEAT 1253 1303 AC repeat.
 CC REPEAT 1318 1330 A repeat (incomplete).
 CC REGION 35 1050 Catalytic (approximate).
 CC REGION 1126 1455 2,4 A, 1 C and 1 AC repeats.
 CC REGION 1126 1455 Glucan-binding (approximate).
 CC VARIANT 21 21 V -> I (in strain GS-5).
 CC VARIANT 81 81 P -> L (in strain MT4239).
 CC VARIANT 106 106 D -> V (in strain GS-5).
 CC VARIANT 116 116 S -> A (in strain GS-5 and strain
 CC MT4467).
 CC VARIANT 126 126 A -> T (in strain GS-5).
 CC VARIANT 150 151 SR -> PK (in strain GS-5, strain MT4239
 CC and strain MT4467).
 CC VARIANT 256 256 A -> V (in strain GS-5 and strain
 CC MT4467).
 CC VARIANT 425 425 R -> N (in strain MT4251).
 CC VARIANT 519 519 Y -> D (in strain MT4245 and strain
 CC MT4251).
 CC VARIANT 538 538 R -> K (in strain MT4245 and strain
 CC MT4251).
 CC VARIANT 545 545 Y -> F (in strain MT4245 and strain
 CC MT4251).
 CC VARIANT 597 597 N -> D (in strain MT4245, strain MT4251,
 CC strain MT4467 and strain MT8148).
 CC VARIANT 600 600 R -> K (in strain MT4245, strain MT4251,

FT VARIANT 601 601 strain MT4467 and strain MT8148).
 FT VARIANT 614 614 A -> T (in strain GS-5).
 FT VARIANT 727 727 M -> T (in strain GS-5).
 FT VARIANT 734 734 T -> I (in strain MT8148).
 FT VARIANT 964 964 A -> V (in strain MT8148).
 FT VARIANT 1113 1113 L -> F (in strain MT4239).
 FT VARIANT 1118 1118 N -> Y (in strain MT4239).
 FT VARIANT 1204 1204 A -> T (in strain MT4239).
 FT VARIANT 1208 1204 I -> V (in strain GS-5, strain MT4239,
 FT strain MT4467 and strain MT8148).
 FT VARIANT 1208 1208 V -> I (in strain MT8148).
 FT VARIANT 1292 1294 DGH -> NGY (in strain GS-5, strain MT4467
 FT and strain MT8148).
 FT VARIANT 1305 1369 Missing (in strain MT4245).
 FT VARIANT 1326 1326 I -> V (in strain MT8148).
 FT VARIANT 1331 1331 T -> A (in strain GS-5, strain MT4239,
 FT strain MT4467 and strain MT8148).
 FT VARIANT 1377 1377 R -> K (in strain MT8148).
 FT VARIANT 1398 1398 V -> I (in strain MT8148).
 FT VARIANT 1424 1424 D -> N (in strain MT4239).
 FT VARIANT 1439 1439 V -> I (in strain MT4239 and strain
 FT MT8148).
 FT VARIANT 1444 1444 S -> P (in strain MT8148).
 FT CONFLICT 1337 1455 QRLYFKNSGVQAGELITERKRIKYDPNSGNEVRNRYVR
 FT TSSGMYNYFGNDGALIGHVVEGRVYVDENGVYRYASHD
 FT ORNHWYDYRRDVGSGSSAVRFRHSRNGFFDNFFRP ->
 FT HASILSLMYRLRESSLSQSVKVSNTMILIPMKFVIM
 FT (in Ref. 1).
 SQ SEQUENCE 1455 AA; 162966 MW; 3CB455A99A4PEC86 CRC64;
 Query Match 40.2%; Score 3183.5; DB 1; Length 1455;
 Best Local Similarity 44.6%; Pred. No. 7,8e-135;
 Matches 691; Conservative 235; Mismatches 445; Indels 179; Gaps 36;
 QY 1 MENKHYLKHVKKQWVTIAVASALATVLGSLVTSSTSSVADETDK--TVTQNSGTT 58
 DB 1 MEKVRFKLRKVKRWVTVSAS-AVVTL---TSLGSLVKADSTDDROQAVTESQ--- 52
 QY 59 ASLVTSPEATKEADKRTNTKEADVLTPAKETNAVETATTTTQATAEAATATTADVAVA 118
 DB 53 ASLVTTSEAKETLTATDTSTATSTATSOPTATVTQNVSTTN-OST---NTTANTANFDVK 108
 QY 119 AVPNKEAVVTTDA-PAVTTTEKAEQ-----PA-----TVKAEVNVTEVKAPE-- 159
 DB 109 PTTTSEQSTQNSDKIIATSKAVNRLTATGKFPANNNTAHSTRTVDKIVPKPKIGLKL 168
 QY 160 --AALKDSEVEAALSILKNIKIDGKYVYNEDGSKHENFAITVNGQLLYFGDKGALTSS 217
 DB 169 QPSSLSQDDIALGNVKNIRKVGKYYKEDGTQLQXVALNINGKTFPFDETGALSNNT 228
 QY 218 TVSFPTPGTTN--IVDGFSSINRAYDSSEASPELIDGYLTADSWYRPASIIKDGVTWQAST 275
 DB 229 LPSKKGNTNNDNTNSFAQYNQVSTDAANFEHVDHYLTAEWSYRPKYLLDKGKTWQST 288
 QY 276 AEDFRPLLMWPNVDVTQVNYLNYMSKVFNLDAKTSSTDKQETLKVAAKDIQIKIEMQITQ 335
 DB 289 EKDFRPLMTWPDQETQRYVYNNYAQLGIHQTTNTATSPQLNLAAQTQTKLEKIT 348
 QY 336 AEKSTQWILRETISAFVKTQPOMNKETENYSKGGEDHLQGGALLVYNDNR--TPWANSYDR 394
 DB 349 AEKNTNWLRTQISAFVKTQSAWNSDSEK----PFDDHLQKGLALYSNNKSLTSQANSNYR 404
 QY 395 RLNRATNTQGTIDKSIILDEQSDPNH-----MGGFDELLANDVDLSNPVQAEQLNQIHY 449
 DB 405 ILNRPTNTQTG-----XKDPRTYATDRTIGGYEFLLANDVDNSNPVQAEQLNLHF 455
 QY 450 LNMWGSIVMGDKDANFDGIRVDADVNDADMLQLTNTYFREYYGNNKSEANLAHSVLE 509
 DB 456 LNMFGNIYANDPDANDFDSIRVDADVNDADLLQIAGDYLKAAKGIKKDKKAANDHLSILE 515
 QY 510 AWSLNDNHNDKTDGAALAMENKORLALLFSIAKPIKERTPAVSPLYNNTP--NTTORDEK 568
 DB 516 AWSYNDTPLYHDDGDGNMINDNRLSLLSYSLAKPLNQRS--GMNPLITNSLVNRTDDNAE 574

```
QY 569 TDWINKSGKAYNEDGTWKQSTIGKNEKYGASGNVYFIRAHNNVDIIAEIIKKEIN 628
Db 575 T-----AAPSIFIRAHDSVQDLIRIIRAEIN 604
QY 629 PKSDGFTITDABMKAOFIYNKMDLSSDKKYTLNINIPAYAVMLQNMETITRVYGDLYT 688
Db 605 PNWGVSTFWEIKAFIYNKDLATEKKYTHYTALSYALLLNKSSVPRVYGDMP 664
QY 689 DDGHYMETSPYDITVINLMKRIKYVSGGQORSYMLPTDGMKNSDVELYRTNEVYS 748
Db 665 DDGQYMAHKTINYEAIETLLKARIKYVSGQAMRNQ-----QVGN-----EIIIS 710
QY 749 VRYGDIIMTANDTESKYSRTSGQVTLVANPKMLDQAKLNVEMKIHAKQKRALIV 808
Db 711 VRYGKALKATDT-GDRITTTSGVAVIEGNNESLRLKASDRVVVNVGAAHKNQAYRPLLL 769
QY 809 GTADGINKFTSDADAIAAGYVKTDSNGVLTGANDIKGYETFDMSGFVAVVWPVGCASN 868
Db 770 TTDNGIKATSHDQ-----AAGLVRYITNDRGELIFTADIKGYANPQVSGYLGWVPVGAAMD 827
QY 869 QDIRVAPSTEAKKEGELTLKATEYDSQLIYEGFSNFQITPDGSDPSVYTNKRKIAENVDL 928
Db 828 QDVRVAASTPSTDGK-SVHQNAALDSRVWPFSGSNFQAF--ATKKEEYTNVVIANKVDK 884
QY 929 FKSNGVTSPEMAPQFVSADGTFDLSVIONGYAPADRYDLAMSKNNKYSGKEDLDALKAKA 988
Db 885 FAEWGVTDPEMAPQVYSSDGSFLDSVIONGYAFTRDYLGIKSPENKYGTTADDLVKAIIKA 944
QY 989 LHKAGIQAATADVPOQIYOLPGKEVVTAATRTDAGRKIADAIIDHSLVYANSKSGKVQ 1048
Db 945 LHSKGILKVNADWPQOMYALPKEVVTAATRVKGYTPVAGSQIKNTLYVVDGKSGSKDQ 1004
QY 1049 KYGGEFLAELKAKYPEMFVKVNMISGKPIDDSVLKQWKABYFNGTNVLRGVGVLSLD 1108
Db 1005 KYGGAFLBELQAKYPELFAKQISTGVPEMDPSVKIKQWSAKYFNGTNILGRGAGVYLD 1064
QY 1109 EATGKYFTVTKGNFIPLQLT---GKEKVIITGFSDDGKGYTFGTSGTQAKSAFVTFNG 1164
Db 1065 QATNTYFSLVSDNTFLPKSLVNPNGHTSSSVTGLVPDGGKGYVYVYSTSGNQAKNAFISLN 1124
QY 1165 NTYYPDARGHMTNSEYSPNGKDVTRFLPGLMGLNFAVIDANGNTYL--YNSKQMYKG 1222
Db 1125 NWTYFDNNGYMTGAQ-SINGANY-FLSNGIQLRNAIY--DNGNKVLSYNGDGRRYEN 1180
QY 1223 GYTKFDVSETDKGKESKVVKFRYFTNEGVMAGVTVIDGFTQYFGEEDGFQAKKLIV-TF 1281
Db 1181 GYTLF-----GQQ-----WRYFQN-GIMAVGLTRIHGAVQYFDASGFQAKGFITTA 1226
QY 1282 KGKTYFDHAHTNGIKDWTW-RNINGKWIYFDANGVAATAQAVINGOKLYFNEGGQVKG 1340
Db 1227 DGKLYRFDKSDGNQISNRNVRNSKGEWFLFDHNGVAVTGTVTFNGQRLYFKPENGVOAKGE 1286
QY 1341 VVKNADGTSYKKEGFGELATNEFFTTDGNVWYAGAKTKVTGAQVINGOHLVFNADGS 1400
Db 1287 FIRDAHGHLRYYPDPSNGNEVRNRFVNSKGEWFLFDHNGIAGVTGRVWNGQRLYFKSNGV 1346
QY 1401 QVKGGVVKNADGTSYKYNATGERLITNEFFTTGDNWYIYGAKSKSVTGEVIGDITYPF 1460
Db 1347 QAKGELITERKRIKYDPNSGNEVRNRYVRITSSGNWYFNGDGYALIG-----1395
QY 1461 AKDGQVKQGTVSAGNRRISYYVSGSGKRAVSTWIEIQGVVYVYFDKXGL 1510
Db 1396 -----W-HVVEGRRVYFDENG 1411
```

RESULT 15

```
Q55265_STRSL PRELIMINARY; PRT; 1577 AA.
ID Q55265_STRSL
AC Q55265;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
```

```
DE Glucosyltransferase precursor.
GN Name=gtfm;
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN (1)
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=95122197; PubMed=7822030;
RA Simpson C.L., Giffard P.M., Jacques N.A.;
RT "Streptococcus salivarius ATCC 25975 possesses at least two genes
coding for primer-independent glucosyltransferases.";
RL Infect. Immun. 63:609-621(1995).
DR EMBL; J35288; AAC41413.1; -; Genomic_DNA.
DR PIR; T30858; T30858.
DR HSSP; P06653; 1H8G.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR004829; Csurface_antigen.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 4.
DR ProDom; PD153432; Ceurface_antigen; 1.
KW Signal; Transferase.
FT SIGNAL 1 38 Potential.
FT CHAIN 39 1577 glucosyltransferase.
SQ SEQUENCE 1577 AA; 175290 MW; 3BFB898A7D3A7BF3 CRC64;
```

Query Match 39.8%; Score 3155; DB 2; Length 1577;

Best Local Similarity 42.5%; Pred. No. 1.7e-133;

Matches 687; Conservative 243; Mismatches 494; Indels 194; Gaps 33;

```
QY 1 MENKHYKXKLVKQWVTIAVASVALATVLGSLVYTSVSADETDQDKT---VTQNSGT 57
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MENKVRFKLHKVKNWVTIGVITLSMALAGSLLAQCKVEADETSAPNGDGLQLSDEG 60
```

```
QY 58 TASLVTSPKATKADKRNTEADVLTPAKETNAVETATTTNTQATAEAT----- 108
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
Db 61 TASLVTTTVTQEQSAQASVASAVATASVSHETSFOAATSQVSEATQATQTSFVASQEVA 120
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
QY 109 -----TATTADVAVAAVPNKEA-----VVTADAPAVTTEKA 139
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
Db 121 VSSQTQSSGQETQTEQVSQGTSTQVAGQTSAQSTPSTEQARPRVLTNAAPATRAA 180
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
QY 140 EEQ-----PATVKAEV-----VNTVKAPE 159
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
Db 181 DSTIRANRNTNITITASGTTPTNVTIITGPNTPKPNVTVTSPTNGTRPNVTIVTQPNQN 240
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
QY 160 AALKDSEV-----EALSLK-----NKNIDGKYYVYVNEGSHKENFAITVNGQLL 205
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
Db 241 KPVQSPQSPKNPVQPNQPSLDYKPVASNLKTDGKYVY-ENGWVKNAAIELDGRLY 299
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
QY 206 YFGKDGALTSSTYSFTPGTTNIVDGFISNNRAYDSSEASPELIDGYLTADSWYRPASII 265
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
Db 300 YFDEGAWDQSKPIYRADALPNNSIYAVYQAYDTSSKFEHLDFLTADSWYRPKQLL 359
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
QY 266 KDGVTWQASTADEDFRPLLMWPNVDTQVNYLNTMSKVFNLDAKYSSTDQKQETLVAARD 325
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
Db 360 KDGKNWTASTEDKYRPELLMTWPDVQVNYLNTMSYQQGFGNKTVTTDMMSYDLAAAABT 419
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
QY 326 IQIKIEQKIQAOKSTQWLRETIASFVKTPQWNKTEYNSKGGEDHLCGGALLVYNDNR 385
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
Db 420 VQRGTEERIGREGNTTLRQLMSDFIKYQPGWNSSED-NLLVVGKDLHQGGALTFLNNSA 478
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
QY 386 TPWANSYRRNLNRTATNOTGT----IDKSIILDEQSDPNHMGGFDFLLANDVLSNPVQA 441
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
Db 479 TSHANSDFRLMNRTPNTQGTGRKYHIDRS-----NGGYELLANDINDNSPAVQA 528
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
QY 442 EQLNQIHYLMWGSIVMGDKDANFDGIRVDVADNVNDADMLQLYTYFRYYGVNKSSEANA 501
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
Db 529 EQLNWLHYIMNIGSILGNDPSANFDGVRIDAVDNVDADLLQIASDYFKKYRVADNEANA 588
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

Qy	502	LAHISVLEAMSLNDHNHNDKTDGALAMENKQRLALLFSLAKPIKERTPAVSPLYNNFTN	561
Db	589	IAHLSIILEAMSYNDHQYNKDTKGQALSIDNPRLRLLTTLFRLKSNYR-GSLERTITNSLN	647
Qy	562	TTORDEKTDWINKDGSKAYNEDGTVKQSTIGKYNEKYGDSAGNTVVFIRAHNNVDIIAE	621
Db	648	NRSS-----QKTPRDANVIFRAHDSEVOAVLAN	678
Qy	622	IIKKEIINPKSDGFTITDAEMKQAFEIYNNKMLSSDKKYTLNNIPAAVAVMLNNMETITRV	681
Db	679	IIKQIINPKTDGFTFTMDLKOAFEIYNADIADAKKKYQYNI PAAVATMLTNKDSITRV	738
Qy	682	YYGDLYTDDGHYMETKSPYDTIYNLMKSRIKYVSGGOAQSRYWLPDGDGKDNDSDVLYR	741
Db	739	YYGDLFTDDGGYMAEKSPYNAIDALLARIKYVAGQ-----DMKVT	782
Qy	742	TN--EVYTSVRYGKDIIMTANDEGSKYSRTSGQVTLVANPNKMLDOSAKLNVEMGKIHA	799
Db	783	LNGYEIMSSVRYGKGAEEANQL-CTAETRNQGMVLVTANRPMKLGANDRLVNMGAHKK	841
Qy	800	NQKTRALIVGTADGINKFTSDADAIAGYVKETDSNGVLTFGANDIKGYEFTFDMSGEVAV	859
Db	842	NQAYRPLLKSTGLATYLDSD--VPAGLVRYTDQGNLTFTADDIAGHSTVEVSGYLA	900
Qy	860	WVPYGASDNQDIRVAPSTEAKKEGELTKATEAVDSOLIVYEGSNSFQTI PDGSDPSPVYTN	919
Db	901	WVPYGASENQDARTKAS--STKGEQVFESAAALDSQVIYEGFSNPODFV--KTPFSQYTN	956
Qy	920	RKIAENYDLFKSMGVTSEFMAPOQVPSADDTGLDSDVIQNGYAFADRVDLAMSNNKYGSK	979
Db	957	RVIAQNAKLFKEWGITSEFAPQVYSSQDGTFLDSIIENGYAPEDRYDIAMSKNNKYGSL	1016
Qy	980	EDLRDALKALHAKIGIAIADWPDDQIIVQLPGKEVWTATRTDAGAKRTIADAIIDHSLYVAN	1039
Db	1017	KDLMADALRALHAEGISAIADWPDDQIYNLPGEVVTASRTNSYGTGPRPNAEINSLYAAK	1076
Qy	1040	SKSGSKDYQAKYGGEPFLAELKAKYPENKFKVMNLTSGRPI DSDSVKLKOWKAEYFNGTNVLE	1099
Db	1077	TRTFGNDPQKYGGAFLDELKAKYPAIFERVQISNGRKLTTNEKITOWSAKIFNGSNIQ	1136
Qy	1100	RGVGYVLSDBATGYFTVTKEGNFIPLQLTGKEKVI TFGSDGKGITYFTGTSQAKSAF	1159
Db	1137	TGARYVLQDNATNQYPSKAGQTFLPKQMT--EITGSGFRVRGDDQYLSIGGYLAKNTF	1194
Qy	1160	VTFNQNT-YYPFDARGHMVTHSEVSPNGKDVYRFLPNCIMLSNAPFYDANGNTYLYNSKG-	1217
Db	1195	IQVGANQWYFYDKXNGNMVTEQWIDGKK--YFPLDNLGLQRLHVLRQSGDGHVYYPDPKV	1252
Qy	1218	QMYKGGYTKPDVSETDKDGKESKVKVPYRPTNEGVMAGVTVJDGFTQYFGE-DGFOAKD	1276
Db	1253	QAFNGFY-----DPAGPRQDV--RYFDGNGQWYRGLHDMYGITTFYDEKTYLQAKD	1301
Qy	1277	KLVTFF-KGKTYYPDAHTGN-GIKDTHWNIGK-WYYPDANGVAATGAQVINGOKLYPNED	1333
Db	1302	KFIRFADGRTRYFIPDTGNLAVNRFAQNPNKAWYLYDSNGYAVTGLQTLINGQYFYDNE	1361
Qy	1334	GSQVKGGVKNADGTSKYKKEGFGELVTNEFFTTDGNVWVYAGANGKTVTGAOVINGOHL	1393
Db	1362	GRQVKGHFVITNNQRY--FLDGSGEIAPSRFVTENNNKWTYYVDGNGKLVKGAQVINGNH	1419
Qy	1394	YFNADGSQVKGGVKNADGTSKYKNASTGBRLTNEFFTTGDNNWYITGANGKSVTGEVKI	1453
Db	1420	YFNNDYSQVKGAW---ANGRY--YDGDGQAVSNQFIQIAANQWYLVNQDGHKVTGLQNT	1474
Qy	1454	GDDTYFFAKDGKQVKGTQVSAGNGRISYYGDSGKRAVSTWIEIQPGVYVYFPDKGLA	1511
Db	1475	NNKYVYFGSNGAQVKGKLLTV-QGKKCYFPAHTGEQVNNRFVBAARGCWYTFYFNSAQQA	1531

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocelebration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 18:56:56 ; Search time 1.99074 Seconds
(without alignments)
4414.224 Million cell updates/sec

Title: US-10-797-821-41

Perfect score: 116
Sequence: 1 GNYWNGGQWAAASAAAAGRY 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*
1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001s.*
5: Geneseq2002s.*
6: Geneseq2003as.*
7: Geneseq2003bs.*
8: Geneseq2004s.*
9: Geneseq2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	116	100.0	20	7	Add93661 Streptoco
2	116	100.0	20	9	Adx37284 Streptoco
3	104	89.7	20	7	Add93639 Streptoco
4	104	89.7	20	9	Adx37282 Streptoco
5	104	89.7	431	7	Add93649 Streptoco
6	104	89.7	431	7	Add93650 Streptoco
7	104	89.7	431	7	Add93653 Streptoco
8	104	89.7	431	9	Adx37272 Streptoco
9	104	89.7	431	9	Adx37273 Streptoco
10	104	89.7	431	9	Adx37276 Streptoco
11	104	89.7	431	9	Aeb91500 Microbial
12	104	89.7	432	7	Add93651 Streptoco
13	104	89.7	432	7	Add93652 Streptoco
14	104	89.7	432	9	Adx37274 Streptoco
15	104	89.7	432	9	Adx37275 Streptoco
16	91	78.4	169	8	Adu69652 Streptoco
17	91	78.4	169	8	Adv88318 Streptoco
18	91	78.4	169	8	Adv81739 Streptoco
19	91	78.4	169	8	Adv79571 Streptoco
20	90	77.6	486	5	Abp66152 Bifidobac
21	89	76.7	132	2	Aaw60944 Streptoco
22	89	76.7	392	6	Abu02747 S. pneumo
23	89	76.7	392	8	Adk47859 Streptoco
24	89	76.7	392	8	Adt50227 S_pneumon

25	89	76.7	392	8	ADT50226	Adt50226 S_pneumon
26	89	76.7	392	8	ADT50165	Adt50165 S_pneumon
27	89	76.7	399	8	ADR4595	Novel S.
28	89	76.7	399	9	AEA58465	Streptoco
29	85	73.3	165	5	ABP29842	Streptoco
30	85	73.3	169	5	ABP27864	Streptoco
31	84	72.4	398	5	ABP25919	Streptoco
32	84	72.4	398	8	ADR3884	S. pyogen
33	84	72.4	398	9	ABP25918	Streptoco
34	83	71.6	395	5	ABP25918	Streptoco
35	83	71.6	447	5	ABP29684	Streptoco
36	83	71.6	447	5	ADU69524	S agalact
37	83	71.6	447	8	ADV88392	Streptoco
38	83	71.6	447	8	ADV81808	Streptoco
39	83	71.6	447	8	ADV79645	Streptoco
40	78	67.2	211	9	ABP91642	Microbial
41	64.5	55.6	318	5	ABP66257	Bifidobac
42	61	52.6	138	5	ABP40216	Streptoco
43	61	52.6	138	8	ADS07279	Staphyloc
44	61	52.6	149	4	AAG82294	S. epider
45	61	52.6	157	4	AAG82888	S. epider

ALIGNMENTS

RESULT 1
ADD93661
ID ADD93661 standard; peptide; 20 AA.
XX
AC ADD93661;
XX
DT 29-JAN-2004 (first entry)
XX
DE Streptococcus mutans glucan binding protein-B peptide fragment.
XX
KW Glucan binding protein-B; GbpB; vaccine; antitoxins; epitope; immunogen.
XX
OS Streptococcus mutans.
XX
PN WO2003075845-A2.
XX
PD 18-SEP-2003.
XX
PF 07-MAR-2003; 2003WO-US006962.
XX
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
XX (FORS-) FORSYTH INST.
PA Smith DJ, Taubman MA;
XX
WPI; 2003-845091/78.
XX
Composition useful as vaccines for dental caries comprises a fragment of a glucan binding protein-B binding to a major histocompatibility complex class II protein.
XX
Claim 11; Page 10; 49pp; English.

The present sequence is that of a peptide fragment of the glucan binding protein-B (GbpB) of Streptococcus mutans strain SK32 ADD93649. The peptide binds to a major histocompatibility complex (MHC) class II protein. It was identified as a potential B cell epitope using a matrix-based algorithm for epitope prediction, which was used to search the primary amino acid sequence of GbpB for known MHC class II binding motifs. The peptide can be used in immunogenic compositions and subunit vaccines for dental caries. These comprise MHC class II protein-binding GbpB peptides covalently linked with peptide subunits (preferably from the catalytic domain) of a glucosyltransferase. The compositions are used in a claimed method of eliciting production of an antibody in a mammal. Diepitopic or multi-epitopic polypeptides can be prepared synthetically or

CC by recombinant DNA technology. Antibodies raised against MHC class II
 CC binding fragments of GbpB can be used in passive immunisation.

XX SQ Sequence 20 AA;

Query Match 100.0%; Score 116; DB 7; Length 20;
 Best Local Similarity 100.0%; Pred. No. 5.9e-08;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNYWNGGQWAASAAAAGRY 20
 |||||
 Db 1 GNYWNGGQWAASAAAAGRY 20

RESULT 2
 ADX37284
 ID ADX37284 standard; peptide; 20 AA.

XX AC ADX37284;

XX DT 21-APR-2005 (first entry)

XX DE Streptococcus mutant glucan binding protein B peptide #29.

XX KW immunogenicity; immune stimulation; glucan binding protein-B;
 KW microparticle; major histocompatibility complex; tooth disease.

XX OS Streptococcus mutans.

XX PN US2005031633-A1.

XX PD 10-FEB-2005.

XX PF 09-MAR-2004; 2004US-00797821.

XX PR 13-APR-1998; 98US-0081550P.

XX PR 08-JAN-1999; 99US-0115142P.

XX PR 12-APR-1999; 99US-00290049.

XX PR 07-MAR-2002; 2002US-0363209P.

XX PR 08-AUG-2002; 2002US-0402483P.

XX PR 07-MAR-2003; 2003US-00383930.

XX PA (SMIT/) SMITH D J.

XX PA (TAUB/) TAUBMAN M A.

XX PI Smith DJ, Taubman MA;

XX PR WPI; 2005-151644/16.

XX PS Claim 4; SEQ ID NO 41; 73pp; English.

XX CC The invention relates to a composition comprising a fragment of a glucan
 CC binding protein-B (GbpB) and a biocompatible microparticle, where the
 CC fragment binds to a major histocompatibility complex (MHC) class II
 CC protein. The composition is useful for producing an antibody for
 CC immunizing mammals against dental caries. This sequence corresponds to a
 CC Streptococcus mutans GbpB-derived peptide of the invention.

XX SQ Sequence 20 AA;

Query Match 100.0%; Score 116; DB 9; Length 20;
 Best Local Similarity 100.0%; Pred. No. 5.9e-08;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNYWNGGQWAASAAAAGRY 20
 |||||
 Db 1 GNYWNGGQWAASAAAAGRY 20

RESULT 3
 ADD93639

XX ID ADD93639 standard; peptide; 20 AA.

XX AC ADD93639;

XX DT 29-JAN-2004 (first entry)

XX DE Streptococcus mutans glucan binding protein-B peptide fragment.

XX KW Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.

XX OS Streptococcus mutans.

XX PN WO2003075845-A2.

XX PD 18-SEP-2003.

XX PF 07-MAR-2003; 2003WO-US006962.

XX PR 07-MAR-2002; 2002US-0363209P.

XX PR 08-AUG-2002; 2002US-0402483P.

XX PA (FORS-) FORSYTH INST.

XX XX Smith DJ, Taubman MA;

XX PI WPI; 2003-845091/78.

XX PS Claim 11; Page 10; 49pp; English.

XX CC The present sequence is that of a peptide comprising amino acid residues
 CC 349-368 of the glucan binding protein-B (GbpB) of Streptococcus mutans
 CC strain SK32 ADD93649. The peptide binds to a major histocompatibility
 CC complex (MHC) class II protein. It was identified as a potential B cell
 CC epitope using a matrix-based algorithm for epitope prediction, which was
 CC used to search the primary amino acid sequence of GbpB for known MHC
 CC class II binding motifs. The peptide can be used in immunogenic
 CC compositions and subunit vaccines for dental caries. These comprise MHC
 CC class II protein-binding GbpB peptides covalently linked with peptide
 CC subunits (preferably from the catalytic domain) of a glucosyltransferase.
 CC The compositions are used in a claimed method of eliciting production of
 CC an antibody in a mammal. Diepitopic or multiepitopic polypeptides can be
 CC prepared synthetically or by recombinant DNA technology. Antibodies
 CC raised against MHC class II binding fragments of GbpB can be used in
 CC passive immunisation.

XX SQ Sequence 20 AA;

Query Match 89.7%; Score 104; DB 7; Length 20;
 Best Local Similarity 100.0%; Pred. No. 2e-06;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNYWNGGQWAASAAAAG 18
 |||||

Db 1 GNYWNGGQWAASAAAAG 18

RESULT 4

ADX37262

XX ID ADX37262 standard; peptide; 20 AA.

XX AC ADX37262;

XX DT 21-APR-2005 (first entry)

XX DE Streptococcus mutant glucan binding protein B peptide #19.

KW immunogenicity; immune stimulation; glucan binding protein-B;
 XX microparticle; major histocompatibility complex; tooth disease.

OS Streptococcus mutans.

PN US2005031633-A1.

XX 10-FEB-2005.

PD 09-MAR-2004; 2004US-00797821.

XX 13-APR-1998; 98US-0081550P.

PR 08-JAN-1999; 99US-0115142P.

PR 12-APR-1999; 99US-00290049.

PR 07-MAR-2002; 2002US-0363209P.

PR 08-AUG-2002; 2002US-0402483P.

PR 07-MAR-2003; 2003US-00383930.

XX (SMIT/) SMITH D J.

PA (TAUB/) TAUBMAN M A.

XX Smith DJ, Taubman MA;

XX WPI; 2005-151644/16.

DR New composition comprising a fragment of a glucan binding protein-B

XX (GbpB) that binds to MHC class II protein, and a biocompatible

PT microparticle, useful for producing an antibody (claimed) for immunizing

PT mammals against dental caries.

XX Claim 4; SEQ ID NO 19; 73pp; English.

PS The invention relates to a composition comprising a fragment of a glucan

XX binding protein-B (GbpB) and a biocompatible microparticle, where the

CC fragment binds to a major histocompatibility complex (MHC) class II

CC protein. The composition is useful for producing an antibody for

CC immunizing mammals against dental caries. This sequence corresponds to a

CC Streptococcus mutans GbpB-derived peptide of the invention.

XX Sequence 20 AA;

SQ

Query Match 89.7%; Score 104; DB 9; Length 20;

Best Local Similarity 100.0%; Pred. No. 2e-06;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GNYWNGGQWAAASAAAG 18

DB 1 GNYWNGGQWAAASAAAG 18

RESULT 5

ADD93649

ID ADD93649 standard; protein; 431 AA.

XX ADD93649;

AC 29-JAN-2004 (first entry)

XX Streptococcus mutans glucan binding protein-B.

XX Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.

XX Streptococcus mutans.

XX Key Location/Qualifiers

FT Region 6..25

FT /note= "HLA-binding peptide"

FT Region 48..67

FT /note= "HLA-binding peptide"

FT Region 52..71

FT /note= "HLA-binding peptide"

FT Region 88..107

FT /note= "HLA-binding peptide"

FT Region 113..132

FT /note= "HLA-binding peptide"

FT Region 117..136

FT /note= "HLA-binding peptide"

FT Region 137..156

FT /note= "HLA-binding peptide"

FT Region 174..193

FT /note= "HLA-binding peptide"

FT Region 194..213

FT /note= "HLA-binding peptide"

FT Region 214..233

FT /note= "HLA-binding peptide"

FT Region 248..267

FT /note= "HLA-binding peptide"

FT Region 289..308

FT /note= "HLA-binding peptide"

FT Region 306..325

FT /note= "HLA-binding peptide"

FT Region 311..330

FT /note= "HLA-binding peptide"

FT Region 349..368

FT /note= "HLA-binding peptide"

FT Region 365..384

FT /note= "HLA-binding peptide"

FT Region 383..402

FT /note= "HLA-binding peptide"

FT Region 403..422

FT /note= "HLA-binding peptide"

XX WO2003075845-A2.

PN 18-SEP-2003.

XX 07-MAR-2003; 2003WO-US0006962.

XX 07-MAR-2002; 2002US-0363209P.

PR 08-AUG-2002; 2002US-0402483P.

XX (FORS-) FORSYTH INST.

XX Smith DJ, Taubman MA;

PI WPI; 2003-845091/78.

XX GENBANK; AY046410.

XX Composition useful as vaccines for dental caries comprises a fragment of

PT a glucan binding protein-B binding to a major histocompatibility complex

PT class II protein.

XX Claim 6; Page 7; 49pp; English.

XX The present sequence is the protein sequence of the glucan binding

CC protein-B (GbpB) of Streptococcus mutans strain SK32. The sequence

CC includes a number of human leucocyte antigen (HLA)-binding epitopes. The

CC invention provides immunogenic compositions and vaccines for dental

CC caries. The compositions comprise major histocompatibility complex (MHC)

CC class II protein-binding GbpB peptides, especially HLA-binding peptides,

CC covalently linked with peptide subunits of a glucosyltransferase. The

CC compositions are used in a claimed method of eliciting production of an

CC antibody in a mammal. Dieptopic or multiepitopic polypeptides can be

CC prepared synthetically or by recombinant DNA technology. Antibodies

CC raised against MHC class II binding fragments of GbpB can be used in

CC passive immunisation.

XX Sequence 431 AA;

SQ Query Match 89.7%; Score 104; DB 7; Length 431;


```
XX PF 09-MAR-2004; 2004US-00797821.
XX XX
PR 13-APR-1998; 98US-0081550P.
PR 08-JAN-1999; 99US-0115142P.
PR 12-APR-1999; 99US-00290049.
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
PR 07-MAR-2003; 2003US-00383930.
XX XX
PA (SMIT/) SMITH D J.
PA (TAUB/) TAUBMAN M A.
XX XX
PI Smith DJ, Taubman MA;
XX WPI; 2005-151644/16.
XX XX
PT New composition comprising a fragment of a glucan binding protein-B
PT (GbpB) that binds to MHC class II protein, and a biocompatible
PT microparticle, useful for producing an antibody (claimed) for immunizing
PT mammals against dental caries.
XX XX
PS Claim 3; SEQ ID NO 29; 73pp; English.
XX XX
PA Smith DJ, Taubman MA;
XX WPI; 2005-151644/16.
XX XX
PT New composition comprising a fragment of a glucan binding protein-B
PT (GbpB) that binds to MHC class II protein, and a biocompatible
PT microparticle, useful for producing an antibody (claimed) for immunizing
PT mammals against dental caries.
XX XX
PS Claim 3; SEQ ID NO 29; 73pp; English.
XX XX
CC The invention relates to a composition comprising a fragment of a glucan
CC binding protein-B (GbpB) and a biocompatible microparticle, where the
CC fragment binds to a major histocompatibility complex (MHC) class II
CC protein. The composition is useful for producing an antibody for
CC immunizing mammals against dental caries. This sequence corresponds to a
CC Streptococcus mutans GbpB protein of the invention.
XX XX
SQ Sequence 431 AA;
Query Match 89.7%; Score 104; DB 9; Length 431;
Best Local Similarity 100.0%; Pred. No. 4.2e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GNYWNGGQWAASAAAAG 18
Db 349 GNYWNGGQWAASAAAAG 366
|||||
RESULT 9
ADX37273
ID ADX37273 standard; protein; 431 AA.
XX AC ADX37273;
XX DT 21-APR-2005 (first entry)
XX DE Streptococcus mutant glucan binding protein B variant #2.
XX KW immunogenicity; immune stimulation; glucan binding protein-B;
XX KW microparticle; major histocompatibility complex; tooth disease.
XX OS Streptococcus mutans.
XX PN US2005031633-A1.
XX PD 10-FEB-2005.
XX PF 09-MAR-2004; 2004US-00797821.
XX PR 13-APR-1998; 98US-0081550P.
XX PR 08-JAN-1999; 99US-0115142P.
XX PR 12-APR-1999; 99US-00290049.
XX PR 07-MAR-2002; 2002US-0363209P.
XX PR 08-AUG-2002; 2002US-0402483P.
XX PR 07-MAR-2003; 2003US-00383930.
XX XX
PA (SMIT/) SMITH D J.
PA (TAUB/) TAUBMAN M A.
XX XX
PI Smith DJ, Taubman MA;
XX WPI; 2005-151644/16.
XX XX
PT New composition comprising a fragment of a glucan binding protein-B
PT (GbpB) that binds to MHC class II protein, and a biocompatible
PT microparticle, useful for producing an antibody (claimed) for immunizing
PT mammals against dental caries.
XX XX
PS Claim 3; SEQ ID NO 33; 73pp; English.
XX XX
CC The invention relates to a composition comprising a fragment of a glucan
CC binding protein-B (GbpB) and a biocompatible microparticle, where the
```

CC fragment binds to a major histocompatibility complex (MHC) class II
 CC protein. The composition is useful for producing an antibody for
 CC immunizing mammals against dental caries. This sequence corresponds to a
 CC Streptococcus mutans GbpB protein of the invention.

SQ Sequence 431 AA;

Query Match 1 89.7%; Score 104; DB 9; Length 431;
 Best Local Similarity 100.0%; Pred. No. 4.2e-05;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNYWNGGOWAASAAAAAG 18
 |||||
 Db 349 GNYWNGGOWAASAAAAAG 366

RESULT 11
 AEB91500
 ID AEB91500 standard; protein; 431 AA.

XX AEB91500;

AC 20-OCT-2005 (first entry)

DT Microbial pathogen adhesin protein sequence, SEQ ID NO:210.

DE algorithm; adhesin; pharmaceutical; vaccine; drug screening;
 KW bordetella pertussis infection; antibacterial; pneumonia;
 KW antinflammatory; respiratory-gen.; gastric ulcer; antiulcer;
 KW gastrointestinal-gen.; urinary tract infection; antimicrobial; uropathic.

XX Streptococcus mutans.

OS WO2005076010-A2.

XX 18-AUG-2005.

XX 07-FEB-2005; 2005WO-IN000037.

XX 06-FEB-2004; 2004IN-DE000173.

PR 20-JUL-2004; 2004US-0589227P.

XX (COUL) COUNCIL SCI & IND RES SOUTH AFRICA.

XX Sachdeva G, Kumar K, Jain P, Brahmachari SK, Ramachandran S;

XX WPI; 2005-597835/61.

XX Computational method for identifying adhesin and adhesin like molecules,
 PT comprises computing sequence-based attributes of protein sequences using
 PT neural network software and training an artificial neural network.

XX Claim 16; SEQ ID NO 210; 402pp; English.

XX The present invention relates to a computational method (M1) for
 CC identifying adhesin and adhesin-like proteins, by computing the sequence-
 CC based attributes of protein sequences using five attribute modules of a
 CC neural network software, training an artificial neural network (ANN) for
 CC each of the computed five attributes, and identifying the adhesin and
 CC adhesin-like proteins having probability of being an adhesin (Pad) as
 CC equal or greater than 0.51. Also claimed is a set of 274 annotated genes
 CC encoding adhesin and adhesin-like proteins, having 274 fully defined 162-
 CC 1151 base pair (SEQ ID NO: 385-658) sequences; a set of 105 hypothetical
 CC genes encoding adhesin and adhesin-like proteins, having 105 fully
 CC defined 306-15876 base pairs (SEQ ID NO: 659-763); a set of 279 annotated
 CC adhesin and adhesin-like proteins, having 279 fully defined 53-3716 base
 CC pair (SEQ ID NO: 1-279); a set of 105 hypothetical adhesin and adhesin-
 CC like proteins, having 105 fully defined 106-5291 base pair (SEQ ID NO:
 CC 280-384) sequences; and a fully connected multilayer feed forward ANN (I)
 CC based on (M1). (M1) is useful for identifying adhesin and adhesin-like
 CC proteins, of therapeutic potential, and identifying and short-listing
 CC proteins for further testing in development of new vaccine formulations
 CC to eliminate diseases caused by various pathogenic organisms. (M1) is

CC useful for identifying putative adhesins that are important in drug
 CC discovery and preventing therapeutics for whooping cough, pneumonia, from
 CC gastric ulcer and urinary tract infections. (M1) identifies adhesins, from
 CC distantly related organisms, and from bacteria belonging to a wide
 CC phylogenetic spectrum. (M1) is capable of predicting adhesive nature of
 CC unique proteins. The present sequence is a microbial pathogen adhesin
 CC protein sequence.

SQ Sequence 431 AA;

Query Match 89.7%; Score 104; DB 9; Length 431;
 Best Local Similarity 100.0%; Pred. No. 4.2e-05;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNYWNGGOWAASAAAAAG 18
 |||||
 Db 349 GNYWNGGOWAASAAAAAG 366

RESULT 12

ADD93651

ID ADD93651 standard; protein; 432 AA.

XX ADD93651;

XX 29-JAN-2004 (first entry)

XX Streptococcus mutans glucan binding protein-B.

XX Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.

XX Streptococcus mutans.

XX WO2003075845-A2.

XX 18-SEP-2003.

XX 07-MAR-2003; 2003WO-US0006962.

XX 07-MAR-2002; 2002US-0363209P.

PR 08-AUG-2002; 2002US-0402483P.

XX (FORS-) FORSYTH INST.

XX Smith DJ, Taubman WA;

XX WPI; 2003-845091/78.

DR GENBANK; AY046412.

XX Composition useful as vaccines for dental caries comprises a fragment of
 PT a glucan binding protein-B binding to a major histocompatibility complex
 PT class II protein.

XX Claim 5; Page 8; 49pp; English.

XX The present sequence is the protein sequence of the glucan binding
 CC protein-B (GbpB) of Streptococcus mutans strain 150P2. The sequence
 CC includes a number of human leucocyte antigen (HLA)-binding epitopes. The
 CC invention provides immunogenic compositions and vaccines for dental
 CC caries. The compositions comprise major histocompatibility complex (MHC)
 CC class II protein-binding GbpB peptides, especially HLA-binding peptides,
 CC covalently linked with peptide subunits of a glucosyltransferase. The
 CC compositions are used in a claimed method of eliciting production of an
 CC antibody in a mammal. Dieptopic or multiepitopic polypeptides can be
 CC prepared synthetically or by recombinant DNA technology. Antibodies
 CC raised against MHC class II binding fragments of GbpB can be used in
 CC passive immunisation.

XX Sequence 432 AA;

Query Match 89.7%; Score 104; DB 7; Length 432;
 Best Local Similarity 100.0%; Pred. No. 4.2e-05;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY      1 GNYWNGGQWAASAAAAG 18
Db      350 GNYWNGGQWAASAAAAG 367

RESULT 13
ADD93652
ID      ADD93652 standard; protein; 432 AA.
XX
AC      ADD93652;
XX
DT      29-JAN-2004 (first entry)
XX
DE      Streptococcus mutans glucan binding protein-B.
XX
KW      Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.
XX
OS      Streptococcus mutans.
XX
PN      WO2003075845-A2.
XX
PD      18-SEP-2003.
XX
PF      07-MAR-2003; 2003WO-US006962.
XX
PR      07-MAR-2002; 2002US-0363209P.
PR      08-AUG-2002; 2002US-0402483P.
XX
PA      (FORS-) FORSYTH INST.
XX
PI      Smith DJ, Taubman MA;
XX
WPI; 2003-845091/78.
DR      GENBANK; AY046413.
XX
CC      Composition useful as vaccines for dental caries comprises a fragment of
PT      a glucan binding protein-B binding to a major histocompatibility complex
PT      class II protein.
XX
PS      Claim 5; Page 8; 49pp; English.
XX
CC      The present sequence is the protein sequence of the glucan binding
CC      protein-B (GbpB) of Streptococcus mutans strain 3SNI. The sequence
CC      includes a number of human leucocyte antigen (HLA)-binding epitopes. The
CC      invention provides immunogenic compositions and vaccines for dental
CC      caries. The compositions comprise major histocompatibility complex (MHC)
CC      class II protein-binding GbpB peptides, especially HLA-binding peptides,
CC      covalently linked with peptide subunits of a glucosyltransferase. The
CC      compositions are used in a claimed method of eliciting production of an
CC      antibody in a mammal. Diepitopic or multiepitopic polypeptides can be
CC      prepared synthetically or by recombinant DNA technology. Antibodies
CC      raised against MHC class II binding fragments of GbpB can be used in
CC      passive immunisation.
XX
SQ      Sequence 432 AA;

Query Match      89.7%; Score 104; DB 7; Length 432;
Best Local Similarity 100.0%; Pred. No. 4.2e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GNYWNGGQWAASAAAAG 18
Db      350 GNYWNGGQWAASAAAAG 367

RESULT 14
ADX37274
ID      ADX37274 standard; protein; 432 AA.
XX
AC      ADX37274;
XX
DT      21-APR-2005 (first entry)
XX

Query Match      89.7%; Score 104; DB 9; Length 432;
Best Local Similarity 100.0%; Pred. No. 4.2e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GNYWNGGQWAASAAAAG 18
Db      350 GNYWNGGQWAASAAAAG 367

RESULT 15
ADX37275
ID      ADX37275 standard; protein; 432 AA.
XX
AC      ADX37275;
XX
DT      21-APR-2005 (first entry)
XX
DE      Streptococcus mutans glucan binding protein B variant #4.
XX
KW      immunogenicity; immune stimulation; glucan binding protein-B;
KW      microparticle; major histocompatibility complex; tooth disease.
XX
OS      Streptococcus mutans.
XX
PN      US2005031633-A1.
XX
PD      10-FEB-2005.
XX
PF      09-MAR-2004; 2004US-00797821.

```

Streptococcus mutant glucan binding protein B variant #3.
immunogenicity; immune stimulation; glucan binding protein-B;
microparticle; major histocompatibility complex; tooth disease.
Streptococcus mutans.
US2005031633-A1.
10-FEB-2005.
09-MAR-2004; 2004US-00797821.
13-APR-1998; 98US-0081550P.
08-JAN-1999; 99US-0115142P.
12-APR-1999; 99US-00290049.
07-MAR-2002; 2002US-0363209P.
08-AUG-2002; 2002US-0402483P.
07-MAR-2003; 2003US-00383930.
(SMIT/) SMITH D J.
(TAUB/) TAUBMAN M A.
Smith DJ, Taubman MA;
WPI; 2005-151644/16.
New composition comprising a fragment of a glucan binding protein-B
(GbpB) that binds to MHC class II protein, and a biocompatible
microparticle, useful for producing an antibody (claimed) for immunizing
mammals against dental caries.
Claim 3; SEQ ID NO 31; 73pp; English.
The invention relates to a composition comprising a fragment of a glucan
binding protein-B (GbpB) and a biocompatible microparticle, where the
fragment binds to a major histocompatibility complex (MHC) class II
protein. The composition is useful for producing an antibody for
immunizing mammals against dental caries. This sequence corresponds to a
Streptococcus mutans GbpB protein of the invention.
Sequence 432 AA;
Query Match 89.7%; Score 104; DB 9; Length 432;
Best Local Similarity 100.0%; Pred. No. 4.2e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GNYWNGGQWAASAAAAG 18
Db 350 GNYWNGGQWAASAAAAG 367
RESULT 15
ADX37275
ID ADX37275 standard; protein; 432 AA.
XX
AC ADX37275;
XX
DT 21-APR-2005 (first entry)
XX
DE Streptococcus mutant glucan binding protein B variant #4.
XX
KW immunogenicity; immune stimulation; glucan binding protein-B;
KW microparticle; major histocompatibility complex; tooth disease.
XX
OS Streptococcus mutans.
XX
PN US2005031633-A1.
XX
PD 10-FEB-2005.
XX
PF 09-MAR-2004; 2004US-00797821.

```
XX 13-APR-1998; 98US-0081550P.
PR 08-JAN-1999; 99US-0115142P.
PR 12-APR-1999; 99US-00290049.
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
PR 07-MAR-2003; 2003US-00383930.
XX (SMIT/) SMITH D J.
PA (TAUB/) TAUBMAN M A.
XX
XX Smith DJ, Taubman MA;
PI
XX WPI; 2005-151644/16.
DR
XX
XX New composition comprising a fragment of a glucan binding protein-B
PT (GbpB) that binds to MHC class II protein, and a biocompatible
PT microparticle, useful for producing an antibody (claimed) for immunizing
PT mammals against dental caries.
XX
XX Claim 3; SEQ ID NO 32; 73pp; English.
XX
XX The invention relates to a composition comprising a fragment of a glucan
CC binding protein-B (GbpB) and a biocompatible microparticle, where the
CC fragment binds to a major histocompatibility complex (MHC) class II
CC protein. The composition is useful for producing an antibody for
CC immunizing mammals against dental caries. This sequence corresponds to a
CC Streptococcus mutans GbpB protein of the invention.
XX
XX Sequence 432 AA;
SQ
Query Match 89.7%; Score 104; DB 9; Length 432;
Best Local Similarity 100.0%; Pred. No. 4.2e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GNYWNGGQWAAASAAAAG 18
Db 350 GNYWNGGQWAAASAAAAG 367
Search completed: February 11, 2006, 19:15:28
Job time : 3.99074 secs
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 19:39:30 ; Search time 0.437886 Seconds
(without alignments)
3776.130 Million cell updates/sec

Title: US-10-797-821-41

Perfect score: 116
Sequence: 1 GNYWNGGQWASAAAAGRY 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/1/iaa/5 COMB.pep.*
2: /cgn2_6/prodata/1/iaa/6 COMB.pep.*
3: /cgn2_6/prodata/1/iaa/H COMB.pep.*
4: /cgn2_6/prodata/1/iaa/PCUTUS COMB.pep.*
5: /cgn2_6/prodata/1/iaa/RE COMB.pep.*
6: /cgn2_6/prodata/1/iaa/backfilese1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	89	76.7	392	US-09-583-110-4374	Sequence 4374, Ap
2	89	76.7	399	US-09-107-433-3230	Sequence 3230, Ap
3	61	52.6	138	US-09-134-001C-5061	Sequence 5061, Ap
4	61	52.6	149	US-09-710-279-1682	Sequence 1682, Ap
5	61	52.6	157	US-09-710-279-2870	Sequence 2870, Ap
6	57	49.1	266	US-09-134-001C-5453	Sequence 5453, Ap
7	56	48.3	257	US-09-710-279-3244	Sequence 3244, Ap
8	56	48.3	264	US-09-134-001C-5035	Sequence 5035, Ap
9	56	48.3	267	US-09-134-001C-4539	Sequence 4539, Ap
10	54.5	47.0	626	US-08-596-300A-7	Sequence 7, Appli
11	54.5	47.0	626	US-08-596-300A-14	Sequence 14, Appli
12	54	46.6	461	US-08-186-222-2	Sequence 2, Appli
13	52	44.8	322	US-09-248-796A-16924	Sequence 16924, A
14	51	44.0	406	US-09-248-796A-20860	Sequence 20860, A
15	50	43.1	113	US-09-513-999C-6331	Sequence 6331, Ap
16	50	43.1	214	US-09-902-540-14258	Sequence 14258, A
17	49	42.2	275	US-09-902-540-16149	Sequence 16149, A
18	49	42.2	339	US-09-252-991A-26841	Sequence 26841, A
19	49	42.2	445	US-09-902-540-13437	Sequence 13437, A
20	48	41.4	28	US-08-425-069-43	Sequence 43, Appli
21	48	41.4	28	US-08-317-844B-43	Sequence 43, Appli
22	48	41.4	624	US-08-756-317-9	Sequence 9, Appli
23	48	41.4	651	US-08-556-978B-19	Sequence 19, Appli
24	48	41.4	651	US-09-247-806-1	Sequence 1, Appli
25	48	41.4	651	US-09-863-859-1	Sequence 1, Appli
26	48	41.4	651	US-09-861-597-1	Sequence 1, Appli
27	48	41.4	718	US-08-425-069-2	Sequence 2, Appli

28	48	41.4	718	1	US-08-317-844B-2	Sequence 2, Appli
29	48	41.4	747	2	US-09-034-177-3	Sequence 3, Appli
30	48	41.4	748	2	US-09-252-991A-32942	Sequence 32942, A
31	47.5	40.9	408	2	US-09-252-991A-31571	Sequence 31571, A
32	47	40.5	146	2	US-08-537-871A-37	Sequence 37, Appli
33	47	40.5	232	2	US-09-252-991A-32746	Sequence 32746, A
34	47	40.5	250	2	US-09-248-796A-22487	Sequence 22487, A
35	46.5	40.1	219	2	US-09-380-015B-2	Sequence 2, Appli
36	46.5	40.1	252	2	US-09-431-887-32	Sequence 32, Appli
37	46.5	40.1	255	1	US-08-242-188-4	Sequence 4, Appli
38	46.5	40.1	255	1	US-08-509-261A-4	Sequence 4, Appli
39	46.5	40.1	255	1	US-08-660-626-10	Sequence 10, Appli
40	46.5	40.1	255	1	US-08-692-892-4	Sequence 4, Appli
41	46.5	40.1	255	1	US-08-713-939A-4	Sequence 4, Appli
42	46.5	40.1	255	1	US-08-868-162A-24	Sequence 24, Appli
43	46.5	40.1	255	2	US-09-031-168-10	Sequence 10, Appli
44	46.5	40.1	255	2	US-09-036-579-4	Sequence 4, Appli
45	46.5	40.1	255	2	US-09-550-374-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-583-110-4374
; Sequence 4374, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: Pneumoniae for Diagnostics and Therapeutics
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 4374
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-4374

Query Match 76.7%; Score 89; DB 2; Length 392;
Best Local Similarity 83.3%; Pred. No. 0.0011;
Matches 15; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GNYWNGGQWASAAAAG 18
|:||||| ||| |||||
Db 304 GDYWGNGAQWATSAAGA 321

RESULT 2
US-107-433-3230
; Sequence 3230, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGN
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354

```
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: <Unknown>
OPERATING SYSTEM: <Unknown>
SOFTWARE: <Unknown>
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,433
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/ 085131
FILING DATE: May 12, 1998
APPLICATION NUMBER: 60/051553
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-011
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 3230:
SEQUENCE CHARACTERISTICS:
LENGTH: 399 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
FEATURE:
NAME/KEY: misc_feature
LOCATION: (B) LOCATION 1...399
SEQUENCE DESCRIPTION: SEQ ID NO: 3230:
US-09-107-433-3230

Query Match 76.7%; Score 89; DB 2; Length 399;
Best Local Similarity 83.3%; Pred. No. 0.0011; 2; Mismatches 1; Gaps 0;
Matches 15; Conservative 1; Indels 2; Indels 0; Gaps 0;

QY 1 GNYWGGGQWASAAAAAG 18
Db 311 GDYWGNGAQWATASAAAG 328

RESULT 3
US-09-134-001C-5061
Sequence 5061, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 5061
LENGTH: 138
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5061

Query Match 52.6%; Score 61; DB 2; Length 138;
Best Local Similarity 55.6%; Pred. No. 1.2;
Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 GNYWGGGQWASAAAAAG 18
Db 82 GSTWGNANSWATAQAAG 99
```

```
RESULT 4
US-09-710-279-1682
Sequence 1682, Application US/09710279
Patent No. 6703492
GENERAL INFORMATION:
APPLICANT: KIMMERLY, WILLIAM JOHN
TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
FILE REFERENCE: PU3480US
CURRENT APPLICATION NUMBER: US/09/710,279
CURRENT FILING DATE: 2000-11-09
PRIOR APPLICATION NUMBER: 60/164,258
PRIOR FILING DATE: 1999-11-09
NUMBER OF SEQ ID NOS: 4472
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1682
LENGTH: 149
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic
OTHER INFORMATION: amino acid sequence
US-09-710-279-1682

Query Match 52.6%; Score 61; DB 2; Length 149;
Best Local Similarity 55.6%; Pred. No. 1.3;
Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 GNYWGGGQWASAAAAAG 18
Db 72 GSTWGNANSWATAQAAG 89

RESULT 5
US-09-710-279-2870
Sequence 2870, Application US/09710279
Patent No. 6703492
GENERAL INFORMATION:
APPLICANT: KIMMERLY, WILLIAM JOHN
TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
FILE REFERENCE: PU3480US
CURRENT APPLICATION NUMBER: US/09/710,279
CURRENT FILING DATE: 2000-11-09
PRIOR APPLICATION NUMBER: 60/164,258
PRIOR FILING DATE: 1999-11-09
NUMBER OF SEQ ID NOS: 4472
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2870
LENGTH: 157
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic
OTHER INFORMATION: amino acid sequence
US-09-710-279-2870

Query Match 52.6%; Score 61; DB 2; Length 157;
Best Local Similarity 55.6%; Pred. No. 1.4;
Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 GNYWGGGQWASAAAAAG 18
Db 72 GSTWGNANSWATAQAAG 89

RESULT 6
US-09-134-001C-5453
Sequence 5453, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
```

```

: TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: GFC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: PRIOR FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 5453
: LENGTH: 286
: TYPE: PRT
: ORGANISM: Staphylococcus epidermidis
: US-09-134-001C-5453

```

```

Query Match      49.1%; Score 57; DB 2; Length 266;
Best Local Similarity 55.6%; Pred. No. 7.6;
Matches 10; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy 1 GNTWNGGQWAAASAAAAG 18
   | | | | | | | |
Db 180 GSLWGNASNWAYSAOAG 197

```

```

RESULT 7
US-09-710-279-3244
; Sequence 3244, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3244
; LENGTH: 257
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-09-710-279-3244

```

Query Match	48.3%	Score 56;	DB 2;	Length 257;
Best Local Similarity	50.0%;	Pred. No.	9.8;	
Matches	9;	Conservative	3;	Mismatches 6; Indels 0; Gaps 0;
Qy	1	GNTWNGGOWAASAAAG	18	
	:	:		
Dd	174	GSTWGNANNWAAAARG	191	

RESULT 8
 US-09-134-001C-5035
 ; Sequence 5035, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: LYNN Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 5035

```
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5035

Query Match      48.3%; Score 56; DB 2; Length 264;
Best Local Similarity 50.0%; Pred.No. 10;
Matches 9; Conservative 3; Mismatches 6; Indels

Qy    1 GNYWNGGGQWAASAAAG 18
       | : |||   || : : : |
Db    181 GSTWGNNANNWNAARSG 198
```

```

RESULT 9
US-09-134-001C-4539
; Sequence 4539, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GFC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4539
; LENGTH: 267
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4539

```

```

Query Match      48.3%; Score 56; DB 2; Length 267;
Best Local Similarity 50.0%; Pred. No. 10;
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 GNYWNGGOWAASAAAAG 18
   | | | | | | | | | |
Db 184 GSTWGNANNWANAARSG 201

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```


TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 626 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
FRAGMENT TYPE:
US-08-596-300A-7

Query Match 47.0%; Score 54.5; DB 1; Length 626;
Best Local Similarity 57.9%; Pred. No. 37;
Matches 11; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

QY 3 YWG-NGGQWAAASAAAGRY 20
Db 323 FWGNSGRTAGKAAAVGRY 341

RESULT 11
US-08-596-300A-14
Sequence 14, Application US/08596300A
Patent No. 5834191
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Production of Heterologous Peptides
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/596,300A
FILING DATE: 13-FEB-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 1321-1-001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 626 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
FRAGMENT TYPE:
US-08-596-300A-14

Query Match 47.0%; Score 54.5; DB 1; Length 626;
Best Local Similarity 57.9%; Pred. No. 37;
Matches 11; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

QY 3 YWG-NGGQWAAASAAAGRY 20
: | | | | | | | | | |

Db 323 FWGNSGRTAGKAAAVGRY 341
RESULT 12
US-08-186-222-2
Sequence 2, Application US/08186222
Patent No. 5559007
GENERAL INFORMATION:
APPLICANT: Suri, Bruno
APPLICANT: Schmitz, Albert
TITLE OF INVENTION: Bacterial Vectors
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/186,222
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/672,205
FILING DATE: 19-MAR-1991
APPLICATION NUMBER: GB 9006400.7
FILING DATE: 22-MAR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Villamizar, JoAnn
REGISTRATION NUMBER: 30,598
REFERENCE/DOCKET NUMBER: 4-17994/A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 785-7121
TELEFAX: (914) 347-5769
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 461 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-186-222-2
Query Match 46.6%; Score 54; DB 1; Length 461;
Best Local Similarity 64.3%; Pred. No. 32;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 5 GNGGQWAAASAAAG 18
: | | | | | | | | | |
Db 376 GNGGQWASNGPQAG 389
RESULT 13
US-09-248-796A-16924
Sequence 16924, Application US/09248796A
Patent No. 6747137
GENERAL INFORMATION:
APPLICANT: Keith Weinstein et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248,796A
CURRENT FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208

; SEQ ID NO 16924
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-16924

Query Match 44.8%; Score 52; DB 2; Length 322;
Best Local Similarity 72.7%; Pred. No. 39;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GNYWNGGQWA 11
|:|||||:
Db 145 GYSGNGGQWS 155

RESULT 14

US-09-248-796A-20860
; Sequence 20860, Application US/09248796A
; Patent No. 6747137

GENERAL INFORMATION:

; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132

; CURRENT APPLICATION NUMBER: US/09/248,796A

; PRIOR FILING DATE: 1999-02-12

; PRIOR APPLICATION NUMBER: US 60/074,725

; PRIOR FILING DATE: 1998-02-13

; PRIOR APPLICATION NUMBER: US 60/096,409

; PRIOR FILING DATE: 1998-08-13

; NUMBER OF SEQ ID NOS: 28208

; SEQ ID NO 20860

; LENGTH: 406

; TYPE: PRT

; ORGANISM: Candida albicans

US-09-248-796A-20860

Query Match 44.0%; Score 51; DB 2; Length 406;
Best Local Similarity 47.4%; Pred. No. 66;
Matches 9; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

QY 1 GNYWNGGQWAASAAAGR 19
|||||:
Db 377 GNYRGNGNWRGNSNRGR 395

RESULT 15

US-09-513-999C-6331
; Sequence 6331, Application US/09513999C
; Patent No. 6783961

GENERAL INFORMATION:

; APPLICANT: Dumas Milne Edwards, J.B.

; APPLICANT: Duclert, A.

; APPLICANT: Giordano, J.Y.

; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.

; Patent No. 6783961

; FILE REFERENCE: 59 US2 REG

; CURRENT APPLICATION NUMBER: US/09/513,999C

; CURRENT FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/122,487

; PRIOR FILING DATE: 1999-02-26

; NUMBER OF SEQ ID NOS: 36681

; SOFTWARE: Patent.pm

; SEQ ID NO 6331

; LENGTH: 113

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: UNSURE

; LOCATION: 24

; OTHER INFORMATION: Xaa=Ala or Ser

; FEATURE:

; NAME/KEY: UNSURE

; LOCATION: 82
; OTHER INFORMATION: Xaa=Ala or Pro
US-09-513-999C-6331

Query Match 43.1%; Score 50; DB 2; Length 113;
Best Local Similarity 55.0%; Pred. No. 24;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 GNYWNGGQWAASAAAGRY 20
|||:
Db 26 GNYFGTHSGAASAAAQY 45

Search completed: February 11, 2006, 19:46:19
Job time : 1.43789 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 20:39:23 ; Search time 1.79012 Seconds
(without alignments)
4668.162 Million cell updates/sec

Title: US-10-797-821-41
Perfect score: 116
Sequence: 1 GNYWNGGQWAAASAAAAGRY 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA_Main:*
1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pep:*
3: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pep:*
4: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pep:*
5: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pep:*
6: /cgn2_6/prodata/1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	116	100.0	20	4	US-10-383-930-41
2	116	100.0	20	5	US-10-797-821-41
3	104	89.7	20	4	US-10-383-930-19
4	104	89.7	20	5	US-10-797-821-19
5	104	89.7	431	4	US-10-383-930-29
6	104	89.7	431	4	US-10-383-930-30
7	104	89.7	431	4	US-10-383-930-33
8	104	89.7	431	5	US-10-797-821-29
9	104	89.7	431	5	US-10-797-821-30
10	104	89.7	431	5	US-10-797-821-33
11	104	89.7	432	4	US-10-383-930-31
12	104	89.7	432	4	US-10-383-930-32
13	104	89.7	432	5	US-10-797-821-31
14	104	89.7	432	5	US-10-797-821-32
15	89	76.7	392	5	US-10-472-928-4652
16	89	76.7	399	5	US-10-617-320-3230
17	84	72.4	398	5	US-10-474-792-600
18	61	52.6	138	4	US-10-724-972A-6574
19	61	52.6	157	5	US-10-470-048B-361
20	61	52.6	166	5	US-10-470-048B-73
21	59	50.9	267	5	US-10-470-048B-74
22	58	50.0	261	5	US-10-470-048B-65
23	57	49.1	266	4	US-10-724-972A-7113
24	57	49.1	360	4	US-10-724-972A-7511
25	56	48.3	77	4	US-10-106-698-6091
26	56	48.3	257	4	US-10-470-048B-362
27	56	48.3	264	4	US-10-724-972A-6539

28	56	48.3	267	4	US-10-724-972A-5110	Sequence 5110, Ap
29	54.5	47.0	647	4	US-10-369-493-3109	Sequence 3109, Ap
30	54	46.6	277	4	US-10-437-963-113400	Sequence 113400, Ap
31	53	45.7	108	6	US-11-097-143-2586	Sequence 2586, Ap
32	52	44.8	101	5	US-10-450-763-57503	Sequence 57503, A
33	52	44.8	363	4	US-10-437-963-170917	Sequence 170917, A
34	52	44.8	1968	4	US-10-369-493-6942	Sequence 6942, Ap
35	52	44.8	1968	4	US-10-369-493-6943	Sequence 6943, Ap
36	51	44.0	113	4	US-10-425-115-218118	Sequence 218118, A
37	51	44.0	375	3	US-09-917-378-3	Sequence 3, Appli
38	51	44.0	375	3	US-09-917-378-6	Sequence 6, Appli
39	51	44.0	762	3	US-09-917-378-1	Sequence 1, Appli
40	51	44.0	1230	4	US-10-437-963-120863	Sequence 120863, A
41	51	44.0	2546	5	US-10-732-923-20517	Sequence 20517, A
42	50	43.1	102	4	US-10-425-115-268735	Sequence 268735, A
43	50	43.1	119	5	US-10-450-763-44127	Sequence 44127, A
44	50	43.1	173	4	US-10-425-115-366795	Sequence 366795, A
45	50	43.1	193	4	US-10-437-963-137412	Sequence 137412, A

ALIGNMENTS

RESULT 1

US-10-383-930-41
; Sequence 41, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 41
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-41

Query Match 100.0%; Score 116; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNYWNGGQWAAASAAAAGRY 20
|||||
Db 1 GNYWNGGQWAAASAAAAGRY 20
|||||

RESULT 2

US-10-797-821-41
; Sequence 41, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049


```
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 30
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-30
```

```
Query Match      89.7%; Score 104; DB 4; Length 431;
Best Local Similarity 100.0%; Pred. No. 6.3e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GNYWNGGQWAAASAAAAG 18
   |||||
Db 349 GNYWNGGQWAAASAAAAG 366
```

RESULT 7

```
US-10-383-930-33
; Sequence 33, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 33
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-33
```

```
Query Match      89.7%; Score 104; DB 4; Length 431;
Best Local Similarity 100.0%; Pred. No. 6.3e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GNYWNGGQWAAASAAAAG 18
   |||||
Db 349 GNYWNGGQWAAASAAAAG 366
```

RESULT 8

```
US-10-797-821-29
; Sequence 29, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
```

```
; SEQ ID NO 29
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-29
```

```
Query Match      89.7%; Score 104; DB 5; Length 431;
Best Local Similarity 100.0%; Pred. No. 6.3e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GNYWNGGQWAAASAAAAG 18
   |||||
Db 349 GNYWNGGQWAAASAAAAG 366
```

RESULT 9

```
US-10-797-821-30
; Sequence 30, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 30
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-30
```

```
Query Match      89.7%; Score 104; DB 5; Length 431;
Best Local Similarity 100.0%; Pred. No. 6.3e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GNYWNGGQWAAASAAAAG 18
   |||||
Db 349 GNYWNGGQWAAASAAAAG 366
```

RESULT 10

```
US-10-797-821-33
; Sequence 33, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
```

```
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 33
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-33
```

```
Query Match      89.7%; Score 104; DB 5; Length 431;
Best Local Similarity 100.0%; Pred. No. 6.3e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 GNYWNGGQWAASAAAAG 18
        |||||
Db      349 GNYWNGGQWAASAAAAG 366
```

```
RESULT 11
US-10-383-930-31
; Sequence 31, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 31
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-31
```

```
Query Match      89.7%; Score 104; DB 4; Length 432;
Best Local Similarity 100.0%; Pred. No. 6.3e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 GNYWNGGQWAASAAAAG 18
        |||||
Db      350 GNYWNGGQWAASAAAAG 367
```

```
RESULT 12
US-10-383-930-32
; Sequence 32, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 32
; LENGTH: 432
```

```
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-32
```

```
Query Match      89.7%; Score 104; DB 4; Length 432;
Best Local Similarity 100.0%; Pred. No. 6.3e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 GNYWNGGQWAASAAAAG 18
        |||||
Db      350 GNYWNGGQWAASAAAAG 367
```

```
RESULT 13
US-10-797-821-31
; Sequence 31, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 31
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-31
```

```
Query Match      89.7%; Score 104; DB 5; Length 432;
Best Local Similarity 100.0%; Pred. No. 6.3e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 GNYWNGGQWAASAAAAG 18
        |||||
Db      350 GNYWNGGQWAASAAAAG 367
```

```
RESULT 14
US-10-797-821-32
; Sequence 32, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
```

; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 32
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-32

Query Match 89.7%; Score 104; DB 5; Length 432;
Best Local Similarity 100.0%; Pred. No. 6.3e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNYWNGGQWAASAAAAAG 18
|:|||||:|||||
Db 350 GNYWNGGQWAASAAAAAG 367

RESULT 15
US-10-472-928-4652
; Sequence 4652, Application US/10472928
; Publication No. US20050020813A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926WO
; CURRENT APPLICATION NUMBER: US/10/472,928
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107658.7
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 4652
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; OTHER INFORMATION: secreted 45 kd protein (usp45)
; OTHER INFORMATION: Cellular location: outside
; OTHER INFORMATION: Similar to strain R6 sequence 15904062 (0.E+01)
US-10-472-928-4652

Query Match 76.7%; Score 89; DB 5; Length 392;
Best Local Similarity 83.3%; Pred. No. 0.0043;
Matches 15; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GNYWNGGQWAASAAAAAG 18
|:|||||:|||||
Db 304 GDYWGNGAQWATSAAAAAG 321

Search completed: February 11, 2006, 20:57:29
Job time : 2.79012 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - proteip search, using sw model

Run on: February 11, 2006, 20:42:47 ; Search time 0.113812 Seconds
(without alignments)
2306.008 Million cell updates/sec

Title: US-10-797-821-41

Perfect score: 116

Sequence: 1 GNYWNGGQWAASAAAAGRY 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 97014 seqs, 1312538 residues

Total number of hits satisfying chosen parameters: 97014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA_New:*
1: /cgn2_6/prodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/prodata/1/pubpaa/US06_NEW_PUB.pep:*
3: /cgn2_6/prodata/1/pubpaa/US07_NEW_PUB.pep:*
4: /cgn2_6/prodata/1/pubpaa/PCT_NEW_PUB.pep:*
5: /cgn2_6/prodata/1/pubpaa/US09_NEW_PUB.pep:*
6: /cgn2_6/prodata/1/pubpaa/US10_NEW_PUB.pep:*
7: /cgn2_6/prodata/1/pubpaa/US11_NEW_PUB.pep:*
8: /cgn2_6/prodata/1/pubpaa/US60_NEW_PUB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	104	89.7	431	7	US-11-052-554A-210
2	84	72.4	398	7	US-11-052-554A-252
3	78	67.2	211	7	US-11-052-554A-352
4	61	52.6	149	6	US-10-793-626-1682
5	61	52.6	157	6	US-10-793-626-2870
6	61	52.6	544	7	US-11-052-554A-358
7	56	48.3	257	6	US-10-793-626-3244
8	49	42.2	224	6	US-10-857-780-25
9	48	41.4	219	6	US-10-517-939-194
10	46.5	40.1	219	6	US-10-867-589-2
11	46	39.7	210	7	US-11-177-509-34
12	46	39.7	211	7	US-11-089-551A-19
13	46	39.7	235	6	US-10-793-626-948
14	46	39.7	283	6	US-10-517-939-180
15	46	39.7	365	7	US-11-052-554A-223
16	46	39.7	618	7	US-11-052-554A-150
17	45	38.8	128	7	US-11-173-071-12
18	45	38.8	128	7	US-11-173-071-14
19	45	38.8	143	6	US-10-793-626-112
20	45	38.8	143	6	US-10-793-626-1940
21	45	38.8	212	7	US-11-089-551A-21
22	45	38.8	294	7	US-11-052-554A-71
23	45	38.8	388	7	US-11-130-821-1
24	45	38.8	1538	7	US-11-052-554A-146
25	44.5	38.4	123	6	US-10-925-366A-217

```

26 44 37.9 141 6 US-10-467-657-748 Sequence 748, App
27 44 37.9 249 7 US-11-054-515-974 Sequence 974, App
28 44 37.9 434 7 US-11-052-554A-167 Sequence 167, App
29 44 37.9 606 7 US-11-052-554A-163 Sequence 163, App
30 44 37.9 1288 7 US-11-052-554A-93 Sequence 93, Appl
31 43.5 37.5 254 7 US-11-054-515-1139 Sequence 1139, App
32 43 37.1 270 6 US-10-485-517-413 Sequence 413, App
33 43 37.1 330 6 US-10-485-517-415 Sequence 415, App
34 43 37.1 1381 7 US-11-052-554A-138 Sequence 138, App
35 43 37.1 1901 7 US-11-052-554A-135 Sequence 135, App
36 42.5 36.6 1660 7 US-11-052-554A-137 Sequence 137, App
37 42 36.2 101 7 US-11-110-424-1 Sequence 1, Appl1
38 42 36.2 101 7 US-11-110-424-2 Sequence 2, Appl1
39 42 36.2 101 7 US-11-110-424-3 Sequence 3, Appl1
40 42 36.2 210 7 US-11-177-509-35 Sequence 35, Appl1
41 42 36.2 210 7 US-11-177-509-36 Sequence 36, Appl1
42 42 36.2 210 7 US-11-177-509-37 Sequence 37, Appl1
43 42 36.2 210 7 US-11-177-509-38 Sequence 38, Appl1
44 42 36.2 210 7 US-11-177-509-39 Sequence 39, Appl1
45 42 36.2 210 7 US-11-177-509-40 Sequence 40, Appl1

```

ALIGNMENTS

```

RESULT 1
US-11-052-554A-210
; Sequence 210, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 210
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans UA159
US-11-052-554A-210

Query Match      89.7%; Score 104; DB 7; Length 431;
Best Local Similarity 100.0%; Pred. No. 1.1e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNYWNGGQWAASAAAAG 18
   |||||
DB 349 GNYWNGGQWAASAAAAG 366

RESULT 2
US-11-052-554A-252
; Sequence 252, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763

```


; Sequence 3244, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: P03480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3244
; LENGTH: 257
; TYPE: PRT
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-3244

Query Match 48.3%; Score 56; DB 6; Length 257;
Best Local Similarity 50.0%; Pred. No. 1.1;
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 GNYWGGQWAAASAAAG 18
|: ||| ||: ||: ||:
Db 174 GSTWGNANNWANAARSQ 191

RESULT 8
US-10-857-780-25
; Sequence 25, Application US/10857780
; Publication No. US20050272043A1
; GENERAL INFORMATION:
; APPLICANT: ROTH, RICHARD B.
; APPLICANT: BRAUN, ANDREAS
; APPLICANT: KAMMERER, STEFAN M.
; APPLICANT: NELSON, MATTHEW ROBERTS
; APPLICANT: RENELAND, RIKARD HENRY
; APPLICANT: HOYAL-WRIGHTSON, CAROLYN R.
; TITLE OF INVENTION: METHODS FOR IDENTIFYING RISK OF BREAST CANCER AND TREATMENTS
; FILE REFERENCE: SEQ-4069-CP
; CURRENT APPLICATION NUMBER: US/10/857,780
; CURRENT FILING DATE: 2004-05-28
; PRIOR APPLICATION NUMBER: 10/723,681
; PRIOR FILING DATE: 2003-11-25
; PRIOR APPLICATION NUMBER: 60/490,234
; PRIOR FILING DATE: 2003-07-24
; PRIOR APPLICATION NUMBER: 60/525,239
; PRIOR FILING DATE: 2003-11-25
; NUMBER OF SEQ ID NOS: 4962
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 25
; LENGTH: 224
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-857-780-25

Query Match 42.2%; Score 49; DB 6; Length 224;
Best Local Similarity 58.3%; Pred. No. 7.4;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 4 WNGGGWAAASAA 15
||| ||: ||: ||:
Db 199 WGGGGKWEATVA 210

RESULT 9
US-10-517-939-194
; Sequence 194, Application US/10517939
; Publication No. US20060003433A1

; GENERAL INFORMATION:
; APPLICANT: Steer, Brian
; APPLICANT: Callen, Walter
; APPLICANT: Healey, Shaun
; APPLICANT: Hazlewood, Geoff
; APPLICANT: Wu, Di
; APPLICANT: Blum, David
; APPLICANT: Esteghlalian, Alireza
; TITLE OF INVENTION: XYLANASES, NUCLEIC ACIDS ENCODING THEM
; AND METHODS FOR MAKING AND USING THEM
; FILE REFERENCE: 564462007901
; CURRENT APPLICATION NUMBER: US/10/517,939
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: PCT/US03/19153
; PRIOR FILING DATE: 2003-06-16
; PRIOR APPLICATION NUMBER: 60/389,299
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 380
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 194
; LENGTH: 279
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
; NAME/KEY: SIGNAL
; LOCATION: (1)...(21)
US-10-517-939-194

Query Match 41.4%; Score 48; DB 6; Length 279;
Best Local Similarity 70.0%; Pred. No. 12;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GNYWNGGQW 10
|: ||| ||: ||:
Db 101 GFYWGNGGKY 110

RESULT 10
US-10-867-589-2
; Sequence 2, Application US/10867589
; Publication No. US20060025575A1
; GENERAL INFORMATION:
; APPLICANT: Carsten Korth
; TITLE OF INVENTION: Immunological Detection of Prions
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kanton Zuerich vertreten durch die Erziehungsadirektion
; STREET: Walchtor
; CITY: Zuerich
; STATE: Zuerich
; COUNTRY: Switzerland
; ZIP: CH-8090
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/867,589
; FILING DATE: 15-Jun-2004
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/380,015
; FILING DATE: 23-Aug-1999
; APPLICATION NUMBER: EP 97102837.8
; FILING DATE: 21-FEB-1997
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 219 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```

; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Bos taurus
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-867-589-2

Query Match      40.1%; Score 46.5; DB 6; Length 219;
Best Local Similarity 39.4%; Pred. No. 15;
Matches 13; Conservative 0; Mismatches 5; Indels 15; Gaps 2;

QY 1 GNYWNGG---GQW-----AASAAAAG 18
Db 75 GGGWGGGGTHGQWNKPSKPTNKLKHVAGAAAAG 107

RESULT 11
US-11-177-509-34
; Sequence 34, Application US/11177509
; Publication No. US20060018918A1
; GENERAL INFORMATION:
; APPLICANT: CHANG, ROWEN J.Y.
; TITLE OF INVENTION: PRODUCTION OF STABILIZED CONFORMATIONAL ISOMERS OF
; FILE REFERENCE: UTHH 1006
; CURRENT APPLICATION NUMBER: US/11/177,509
; CURRENT FILING DATE: 2005-07-08
; PRIOR APPLICATION NUMBER: 10/210,862
; PRIOR FILING DATE: 2002-08-01
; PRIOR APPLICATION NUMBER: 10/025,976
; PRIOR FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/258,576
; PRIOR FILING DATE: 2000-12-27
; NUMBER OF SEQ ID NOS: 88
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 34
; LENGTH: 210
; TYPE: PRT
; ORGANISM: Mus musculus
US-11-177-509-34

Query Match      39.7%; Score 46; DB 7; Length 210;
Best Local Similarity 38.2%; Pred. No. 17;
Matches 13; Conservative 0; Mismatches 5; Indels 16; Gaps 2;

QY 1 GNYWNGG---GQW-----AASAAAAG 18
Db 64 GGGWGGGGTHGQWNKPSKPTNKLKHVAGAAAAG 97

RESULT 12
US-11-089-551A-19
; Sequence 19, Application US/11089551A
; Publication No. US20050266242A1
; GENERAL INFORMATION:
; APPLICANT: Lindquist et al.
; TITLE OF INVENTION: ELECTRICAL CONDUCTORS AND DEVICES FROM PRION-LIKE PROTEINS
; FILE REFERENCE: 30554/40025A
; CURRENT APPLICATION NUMBER: US/11/089,551A
; CURRENT FILING DATE: 2005-03-24
; PRIOR APPLICATION NUMBER: US 60/559,286
; PRIOR FILING DATE: 2004-03-31
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 211
; TYPE: PRT
; ORGANISM: MOUSE
US-11-089-551A-19

Query Match      39.7%; Score 46; DB 7; Length 211;
Best Local Similarity 38.2%; Pred. No. 17;
Matches 13; Conservative 0; Mismatches 5; Indels 16; Gaps 2;

```

```

Matches 13; Conservative 0; Mismatches 5; Indels 16; Gaps 2;

QY 1 GNYWNGG---GQW-----AASAAAAG 18
Db 65 GGGWGGGGTHGQWNKPSKPTNKLKHVAGAAAAG 98

RESULT 13
US-10-793-626-948
; Sequence 948, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS34800S
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 948
; LENGTH: 235
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-10-793-626-948

Query Match      39.7%; Score 46; DB 6; Length 235;
Best Local Similarity 47.1%; Pred. No. 19;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 NYWNGGQWAAASAAAAG 18
Db 72 NFTSNGSEWSYAVAG 88

RESULT 14
US-10-517-939-180
; Sequence 180, Application US/10517939
; Publication No. US20060003433A1
; GENERAL INFORMATION:
; APPLICANT: Steer, Brian
; APPLICANT: Callen, Walter
; APPLICANT: Healey, Shaun
; APPLICANT: Hazlewood, Geoff
; APPLICANT: Wu, Di
; APPLICANT: Blum, David
; APPLICANT: Esteghalian, Alireza
; TITLE OF INVENTION: XYLANASES, NUCLEIC ACIDS ENCODING THEM
; FILE REFERENCE: 564462007901
; CURRENT APPLICATION NUMBER: US/10/517,939
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: PCT/US03/19153
; PRIOR FILING DATE: 2003-06-16
; PRIOR APPLICATION NUMBER: 60/389,299
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 380
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 180
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
; NAME/KEY: SIGNAL
; LOCATION: (1)...(19)
US-10-517-939-180

```

Query Match 39.7%; Score 46; DB 6; Length 283;
Best Local Similarity 70.0%; Pred. No. 22;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GNYWNGGQW 10
DB 104 GFWGNGGQF 113

RESULT 15

US-11-052-554A-223
; Sequence 223, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 223
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Streptococcus mutans UA159
US-11-052-554A-223

Query Match 39.7%; Score 46; DB 7; Length 365;
Best Local Similarity 87.5%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 YWNGGGQW 10
DB 278 YWNGGGQW 285

Search completed: February 11, 2006, 20:59:02
Job time : 1.11381 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 19:15:50 ; Search time 0.412809 Seconds
(without alignments)
4661.567 Million cell updates/sec

Title: US-10-797-821-41

Perfect score: 116

Sequence: 1 GNYWNGGQWAASAAAAGRY 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 80:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	89	76.7	392	2	G95258
2	89	76.7	392	2	B98124
3	61	52.6	166	2	C90029
4	59	50.9	267	2	F90028
5	58	50.0	255	2	G90061
6	57	49.1	358	2	H36891
7	57	49.1	475	2	C86863
8	54.5	47.0	626	2	T49625
9	54.5	47.0	626	2	S36364
10	54	46.6	355	2	E95843
11	54	46.6	461	2	JN0097
12	53	45.7	161	2	E85354
13	53	45.7	456	2	E86903
14	52.5	45.3	71	2	S45608
15	52	44.8	732	2	D64754
16	52	44.8	732	2	E85518
17	52	44.8	732	2	B90668
18	52	44.8	1968	1	S05697
19	51	44.0	627	2	F95867
20	50	43.1	279	2	D89848
21	50	43.1	460	2	G82847
22	50	43.1	611	2	S60040
23	50	43.1	1052	2	T14343
24	49.5	42.7	387	2	T52451
25	49.5	42.7	2639	2	T31328
26	49	42.2	187	2	G83047
27	49	42.2	205	2	G81055
28	49	42.2	283	2	G82860
29	49	42.2	502	2	B70845

30 48 41.4 257 2 A27872 outer membrane pro
31 48 41.4 562 2 B70953 hypothetical glyci
32 48 41.4 624 2 JC6023 poly(3-hydroxyalka
33 48 41.4 718 2 A36068 major ampullate fi
34 47 40.5 59 2 T43106 hypothetical prote
35 47 40.5 145 2 A87574 hypothetical prote
36 47 40.5 435 2 T15143 hypothetical prote
37 47 40.5 1246 2 G90887 cryptic nitrate re
38 47 40.5 1246 2 B85730 cryptic nitrate re
39 46.5 40.1 141 2 I47178 Ig heavy chain var
40 46.5 40.1 252 2 JC6175 prion protein - ra
41 46.5 40.1 256 2 JU0268 major prion protei
42 46.5 40.1 256 2 S37149 prion protein - go
43 46.5 40.1 256 2 A54281 major prion protei
44 46.5 40.1 264 2 A54330 major prion protei
45 46.5 40.1 264 2 S37137 prion protein - gr

ALIGNMENTS

RESULT 1

G95258 secreted 45 kd protein [imported] - Streptococcus pneumoniae (strain TIGR4)

C:Species: Streptococcus pneumoniae

C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 09-Jul-2004

C:Accession: G95258

R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hei-

non, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,

non, T.; Hickey, E.K.; Holt, I.E.

Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,

A>Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.

A:Reference number: A95000; MUID:21357209; PMID:11463916

A:Accession: G95258

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-392 <KUR>

A:Cross-references: UNIPROT:Q97N55; UNIPROT:Q8DMV4; UNIPARC:UPI0000051B81; GB:AE005672;

A:Experimental source: strain TIGR4

C:Genetics:

A:Gene: SP2216

Query Match 76.7%; Score 89; DB 2; Length 392;
Best Local Similarity 83.3%; Pred No. 0.0002;
Matches 15; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GNYWNGGQWAASAAAAG 18

Db 304 GDYWGNGAQWATSAAGA 321

:||||| ||| |||||

:||||| ||| |||||

:||||| ||| |||||

:||||| ||| |||||

:||||| ||| |||||

:||||| ||| |||||

:||||| ||| |||||

:||||| ||| |||||

:||||| ||| |||||

:||||| ||| |||||

:||||| ||| |||||

:||||| ||| |||||

:||||| ||| |||||

:||||| ||| |||||

:||||| ||| |||||

:||||| ||| |||||

:||||| ||| |||||

:||||| ||| |||||

:||||| ||| |||||

:||||| ||| |||||

:||||| ||| |||||

:||||| ||| |||||

:||||| ||| |||||

:||||| ||| |||||

:||||| ||| |||||

:||||| ||| |||||

:||||| ||| |||||

:||||| ||| |||||


```
Query Match          76.7%;   Score 89;   DB 2;   Length 392;
Best Local Similarity 83.3%;   Pred. No. 0.0002; 2;   Indels 0;
Matches 15; Conservative 1; Mismatches 2; Gaps 0;

QY 1 GNYWNGGQWAASAAAAG 18
   |:|||||:|||||
Db 304 GDYWGNGAQWATSAAAAG 321

RESULT 3
C90029
hypothetical protein SA2097 [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: C90029
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayaashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: C90029
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-166 <KUR>
A:Cross-references: UNIPROT:Q99RW9; UNIPARC:UPI000005434E; GB:BA000018; PID:g13702104; F
A:Experimental source: strain N315
C:Genetics:
A:Gene: SA2097

Query Match          52.6%;   Score 61;   DB 2;   Length 166;
Best Local Similarity 55.6%;   Pred. No. 0.44; 6;   Indels 0;
Matches 10; Conservative 2; Mismatches 6; Gaps 0;

QY 1 GNYWNGGQWAASAAAAG 18
   |:|||||:|||||
Db 82 GSTWGNWNAWAAQAAG 99

RESULT 4
F90028
hypothetical protein saaA [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: F90028
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayaashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: F90028
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-267 <KUR>
A:Cross-references: UNIPROT:Q99RX4; UNIPARC:UPI000005225C; GB:BA000018; PID:g13702099; F
A:Experimental source: strain N315
C:Genetics:
A:Gene: saaA

Query Match          50.9%;   Score 59;   DB 2;   Length 267;
Best Local Similarity 55.0%;   Pred. No. 1.3; 2;   Indels 0;
Matches 10; Conservative 2; Mismatches 6; Gaps 0;

QY 1 GNYWNGGQWAASAAAAG 18
   |:|||||:|||||
Db 184 GSTWGNWNAWAAARAG 201

RESULT 5
G90061
hypothetical protein SA2353 [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
```

```
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: G90061
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayaashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: G90061
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-255 <KUR>
A:Cross-references: UNIPROT:Q53587; UNIPARC:UPI00000D7806; GB:BA000018; PID:g13702516; F
A:Experimental source: strain N315
C:Genetics:
A:Gene: SA2353

Query Match          50.0%;   Score 58;   DB 2;   Length 255;
Best Local Similarity 50.0%;   Pred. No. 1.6; 4;   Indels 0;
Matches 9; Conservative 4; Mismatches 5; Gaps 0;

QY 1 GNYWNGGQWAASAAAAG 18
   |:|||||:|||||
Db 172 GSTWGNWNAWAAASSG 189

RESULT 6
H36891
transfer complex protein trsg - Staphylococcus aureus
C:Species: Staphylococcus aureus
C:Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 09-Jul-2004
C:Accession: H36891
R:Morton, T.M.; Eaton, D.M.; Johnston, J.L.; Archer, G.L.
J. Bacteriol. 175, 4436-4447, 1993
A:Title: DNA sequence and units of transcription of the conjugative transfer gene comple
A:Reference number: A36891; MUID:93322322; PMID:7687249
A:Accession: H36891
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-358 <MOR>
A:Cross-references: UNIPROT:Q07727; UNIPARC:UPI00000B98DF; GB:111998; NID:g310606; PIDN
C:Genetics:
A:Gene: trsg
C:Superfamily: Staphylococcus aureus transfer complex protein trsg

Query Match          49.1%;   Score 57;   DB 2;   Length 358;
Best Local Similarity 60.0%;   Pred. No. 3; 9;   Indels 0;
Matches 9; Conservative 1; Mismatches 5; Gaps 0;

QY 4 WNGGQWAASAAAAG 18
   |:|||||:|||||
Db 271 WNGGGDWGDNKAQAG 285

RESULT 7
C86863
N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) [imported] - Lactococcus lactis subsp.
N:Alternate names: N-acetylmuramidase
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 09-Jul-2004
C:Accession: C86863
R:Boletín, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: C86863
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-475 <STO>
A:Cross-references: UNIPROT:Q9CED5; UNIPARC:UPI00000C6B58; GB:AE005176; PID:g12724943; F
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: acmb
```

C;Keywords: hydrolase

Query Match 49.1%; Score 57; DB 2; Length 475;
Best Local Similarity 45.0%; Pred. No. 3.9;
Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 1 GNYWNGGQWAASAAAGRY 20
| : ||||| : ||||| :
Db 385 GTHMGNGGNGINAQAQGYF 404

RESULT 8
T49625
Glucan 1,4-alpha-glucosidase [imported] - Neurospora crassa
N;Alternate names: protein B5022.70
C;Species: Neurospora crassa
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2003
C;Accession: T49625
R;Schulte, U.; Aign, V.; Hohelsel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura, submitted to the Protein Sequence Database, May 2000
A;Reference number: Z25022
A;Accession: T49625
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-626 <SCH>
A;Cross-references: UNIPARC:UPI000006A8CB; EMBL:AL355932; GSPDB:GN001116; NCSP:B5022.70
A;Experimental source: BAC clone B5022; strain OR74A
C;Genetics:
A;Gene: NCSP:B5022.70
A;Map position: 6 |
A;Introns: 82/1
C;Superfamily: glucan 1,4-alpha-glucosidase with starch-binding domain; glucan 1,4-alpha

Query Match 47.0%; Score 54.5; DB 2; Length 626;
Best Local Similarity 57.9%; Pred. No. 11;
Matches 11; Conservative 2; Mismatches 5; Indels 1; Gaps 1;
QY 3 YWG-NGGQWAASAAAGRY 20
| : ||||| : ||||| :
Db 323 FWGVNSGRTAGKAAAVGRY 341

RESULT 9
S36364
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) precursor - Neurospora crassa
N;Alternate names: glucoamylase; glucoamylase
C;Species: Neurospora crassa
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 02-Jun-2003
C;Accession: S36364; S13710; S13711; S25539
R;Stone, P.J.; Makoff, A.J.; Parish, J.H.; Radford, A.
Curr. Genet. 24, 205-211, 1993
A;Title: Cloning and sequence analysis of the glucoamylase gene of Neurospora crassa.
A;Reference number: S36364; MUID:94037144; PMID:8221928
A;Accession: S36364
A;Molecule type: DNA
A;Residues: 1-626 <STO>
A;Cross-references: UNIPARC:UPI00014A557; EMBL:X67291
R;Koh-Luar, S.I.; Parish, J.H.; Bleasby, A.J.; Pappin, D.J.C.; Ainley, K.; Johansen, F.H.
Enzyme Microb. Technol. 11, 692-695, 1989
A;Title: Exported proteins of Neurospora crassa: 1-glucoamylase.
A;Reference number: S13710

A;Accession: S13710
A;Molecule type: protein
A;Residues: 36-60,'X',62,'X',64-65 <KOH>
A;Cross-references: UNIPARC:UPI0000175A5C
R;Koh-Luar, S.I.; Parish, J.H.; Bleasby, A.J.; Pappin, D.J.C.; Ainley, K.; Johansen, F.H.
submitted to the Protein Sequence Database, January 1990
A;Description: Exported proteins of Neurospora crassa 1: - glucoamylase.
A;Reference number: S13711
A;Accession: S13711
A;Molecule type: protein
A;Residues: 36-60,'X',62,'X',64-65 <KOW>
A;Cross-references: UNIPARC:UPI0000175A5C

C;Genetics:

A;Gene: Gla-1
A;Introns: 82/2
C;Superfamily: glucan 1,4-alpha-glucosidase with starch-binding domain; glucan 1,4-alpha
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-35/Domain: propeptide #status predicted <PRO>
F;33-458/Domain: glucan 1,4-alpha-glucosidase homology <GAG>
F;36-626/Product: glucan 1,4-alpha-glucosidase #status experimental <MAT>

Query Match 47.0%; Score 54.5; DB 2; Length 626;
Best Local Similarity 57.9%; Pred. No. 11;
Matches 11; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

QY 3 YWG-NGGQWAASAAAGRY 20
| : ||||| : ||||| :
Db 323 FWGVNSGRTAGKAAAVGRY 341

RESULT 10
E95843
hypothetical protein SMB20010 [imported] - Sinorhizobium meliloti (strain 1021) magapla
C;Species: Sinorhizobium meliloti
C;Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 09-Jul-2004
C;Accession: E95843
R;Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Herna
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A;Title: The complete sequence of the 1.683-kb pSymb megaplasmid from the N2-fixing end
A;Reference number: A95842; MUID:21396508; PMID:11481431
A;Accession: E95843
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-355 <KUR>
A;Cross-references: UNIPROT:Q92XE9; UNIPARC:UPI00000CB3A1; GB:AL591985; PIDN:CAC48413.1
A;Experimental source: strain 1021, megaplasmid pSymb
R;Galberet, F.; Finan, T.M.; Dong, S.R.; Punler, A.; Abola, P.; Ampe, F.; Barloy-Hubler
peia, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yen, K.
A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A;Reference number: A96039; MUID:21368234; PMID:11474104
A;Contents: annotation
C;Genetics:
C;Gene: SMB20010
A;Genome: plasmid

Query Match 46.6%; Score 54; DB 2; Length 355;
Best Local Similarity 69.2%; Pred. No. 7.4;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 WNGGQWAASAAA 16
| : ||||| : ||||| :
Db 151 WATGGQWAAAFAA 163

RESULT 11
JN0097
secreted 45K protein precursor - Lactococcus lactis
C;Species: Lactococcus lactis
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999
C;Accession: JN0097
R;van Asseldonk, M.; Rutten, G.; Oteman, M.; Sieren, R.J.; de Vos, W.M.; Simons, G.
Gene 95, 155-160, 1990
A;Title: Cloning of usp45, a gene encoding a secreted protein from Lactococcus lactis
A;Reference number: JN0097; MUID:91071599; PMID:2123812
A;Accession: JN0097
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-461 <VAN>
A;Cross-references: UNIPARC:UPI000016D72B; GB:M35374

Query Match 46.6%; Score 54; DB 2; Length 461;
 Best Local Similarity 64.3%; Pred. No. 9.5;
 Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 GNGGQWAAASAAAAG 18
 |||||:::
 Db 376 GNGGQWASNGPAQG 389

RESULT 12

E85354
 C:Species: Arabidopsis thaliana [imported] - Arabidopsis thaliana
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
 C:Accession: E85354
 R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring
 Nature 402, 769-777, 1999
 A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
 A:Reference number: A85001; MUID:20083488; PMID:10617198
 A:Accession: E85354
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-161 <STO>
 A:Cross-references: UNIPROT:Q9M0C8; UNIPARC:UPI00000A03A9; GB:NC_001268; NID:g7269932; E
 C:Genetics:
 A:Gene: At4g30320
 A:Map position: 4
 C:Superfamily: pathogenesis-related leaf protein

Query Match 45.7%; Score 53; DB 2; Length 161;
 Best Local Similarity 50.0%; Pred. No. 4.8;
 Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 3 YMGNGQWAAASAAAAG 18
 :|||:|:|:
 Db 81 FWGSGNRWGPSQAAYG 96

RESULT 13

E86903
 C:Species: Lactococcus lactis subsp. lactis (strain IL1403)
 C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 09-Jul-2004
 C:Accession: E86903
 R:Solotkin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarre, K.; Weissenbach, J.; Ehrlich
 Genome Res. 11, 731-753, 2001
 A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s
 A:Reference number: A86625; MUID:21235186; PMID:11337471
 A:Accession: E86903
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-456 <STO>
 A:Cross-references: UNIPROT:Q9CDJ1; UNIPARC:UPI00000C6BFC; GB:AE005176; PID:gl2725296; E
 A:Experimental source: strain IL1403
 C:Genetics:
 A:Gene: usp45

Query Match 45.7%; Score 53; DB 2; Length 456;
 Best Local Similarity 64.3%; Pred. No. 13;
 Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 5 GNGGQWAAASAAAAG 18
 |||||:::
 Db 371 GNGGQWATNGPAQG 384

RESULT 14

S45608
 C:Species: Rhodocyclus gelatinosus
 N:Alternate names: peripheral antenna complex B800-850 alpha chain
 C:Date: 10-Dec-1994 #sequence_revision 19-Jul-1996 #text_change 09-Jul-2004
 C:Accession: S45608; S66177

R:Brumisholz, R.A.; Suter, F.; Zuber, H.

Eur. J. Biochem. 222, 667-675, 1994

A:Title: Structural and spectral characterisation of the antenna complexes of Rhodocycl

A:Reference number: S45608; MUID:94291666; PMID:8020505

A:Accession: S45608

A:Molecule type: protein

A:Residues: 1-71 <BRU>

A:Cross-references: UNIPROT:P77799; UNIPARC:UPI00000B40B7

A:Experimental source: DSM 149

A:Accession: S66177

A:Molecule type: protein

A:Residues: 1-71 <BR2>

A:Cross-references: UNIPARC:UPI00000B40B7

A:Experimental source: DSM 151

C:Superfamily: light-harvesting protein alpha chain

C:Keywords: antenna complex; bacteriochlorophyll; light-harvesting polypeptide; membran

Query Match 45.3%; Score 52.5; DB 2; Length 71;

Best Local Similarity 70.6%; Pred. No. 2.6;

Matches 12; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

QY 1 GNYWNGGQWAAASAAA 17
 |||||:|:
 Db 42 GAYW-NGGKKAASAAA 57

RESULT 15

D64754
 C:Species: Escherichia coli
 C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 09-Jul-2004
 C:Accession: D64754
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C
 A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: D64754

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-732 <BLAT>

A:Cross-references: UNIPROT:P77489; UNIPARC:UPI000013A0B1; GB:AE000136; GB:U000096; NID:

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: yagR

C:Superfamily: carbon monoxide dehydrogenase molybdoprotein

C:Keywords: oxidoreductase

Query Match 44.8%; Score 52; DB 2; Length 732;

Best Local Similarity 50.0%; Pred. No. 27;

Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 5 GNGGQWAAASAAAAGRY 20
 |:|||||:|:
 Db 505 GSGGQWAGTSTSGVY 520

Search completed: February 11, 2006, 19:42:58

Job time : 1.41281 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.
OM protein - protein search, using sw model
Run on: February 11, 2006, 18:57:35 ; Search time 2.69869 Seconds
(without alignments)
5228.676 Million cell updates/sec
Title: US-10-797-821-41
Perfect score: 116
Sequence: 1 GNYWNGGQWAAASAAAAGRY 20
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 2166443 seqs, 705528306 residues
Total number of hits satisfying chosen parameters: 2166443
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : Uniprot 05.80.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	104	89.7	431	2 Q938V0_STRMU	Q938V0 streptococc
2	104	89.7	431	2 Q938V3_STRMU	Q938V3 streptococc
3	104	89.7	431	2 Q9AG98_STRMU	Q9AG98 streptococc
4	104	89.7	431	2 Q8DMW3_STRMU	Q8DMW3 streptococc
5	104	89.7	432	2 Q938V1_STRMU	Q938V1 streptococc
6	104	89.7	432	2 Q938V2_STRMU	Q938V2 streptococc
7	91	78.4	169	2 Q8E3F4_STRAJ	Q8E3F4 streptococc
8	90	77.6	486	2 Q8G4C1_BIFLO	Q8G4C1 bifidobacte
9	89	76.7	392	2 Q8DMY4_STRR6	Q8DMY4 streptococc
10	89	76.7	392	2 Q97N55_STRPN	Q97N55 streptococc
11	85	73.3	169	2 Q8DXT4_STRAS	Q8DXT4 streptococc
12	84	72.4	125	2 Q8RQM2_STRRT	Q8RQM2 streptococc
13	84	72.4	398	2 Q5XEL1_STRP6	Q5XEL1 streptococc
14	84	72.4	398	2 Q9A1Z8_STRPY	Q9A1Z8 streptococc
15	84	72.4	398	2 Q7CNCQ7_STRP8	Q7CNCQ7 streptococc
16	84	72.4	398	2 Q8P318_STRP3	Q8P318 streptococc
17	83	71.6	447	2 Q9AKA4_STRAG	Q9AKA4 streptococc
18	83	71.6	447	2 Q8EZHL_STRAS	Q8EZHL streptococc
19	83	71.6	447	2 Q8E7X9_STRAJ	Q8E7X9 streptococc
20	80	69.0	132	2 Q8RQM3_9STRE	Q8RQM3 streptococc
21	78	67.2	125	2 Q8RQM1_STRMU	Q8RQM1 streptococc
22	78	67.2	211	2 Q8DVU8_STRMU	Q8DVU8 streptococc
23	76	65.5	128	2 Q8RQM0_9STRE	Q8RQM0 streptococc
24	75	64.7	461	2 Q56SA7_STRTR	Q56SA7 streptococc
25	67	57.8	263	2 Q54487_STACA	Q54487 streptococc
26	66	56.9	129	2 Q5M137_STRT1	Q5M137 streptococc
27	66	56.9	474	2 Q5M6K4_STRT2	Q5M6K4 streptococc
28	66	56.9	482	2 Q5M5M6_STRT2	Q5M5M6 streptococc
29	66	56.9	485	2 Q5M212_STRT1	Q5M212 streptococc
30	65	56.0	286	2 Q4L410_STAHJ	Q4L410 staphylococ
31	65	56.0	544	2 Q840V8_STRMU	Q840V8 streptococc

RESULT 1
Q938V0_STRMU
ID Q938V0_STRMU PRELIMINARY; PRT; 431 AA.
AC Q938V0_STRMU PRELIMINARY; PRT; 431 AA.
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Glucan-binding protein B.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
ON NCBI_TaxID=1309;
RX [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=SSM3;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gbpB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224 (2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=SSM3;
RX MEDLINE=21481971; PubMed=11598068;
RA Mattos-Grauer R.O., Jin S., King W.F., Chen T., Smith D.J.,
Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding
protein B and analysis of genetic diversity and protein production in
clinical isolates.";
RL Infect. Immun. 69:6931-6941 (2001).
DR EMBL; AY046414; AAK94504.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIBA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PDS0911; CHAP; 1.
SQ SEQUENCE 431 AA; 44650 MW; 05D38D8D8BC4609F CRC64;
Query Match 89.7%; Score 104; DB 2; Length 431;
Best Local Similarity 100.0%; Pred. No. 1.6e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GNYWNGGQWAAASAAAAG 18
|||||
Db 349 GNYWNGGQWAAASAAAAG 366
RESULT 2
Q938V3_STRMU
ID Q938V3_STRMU PRELIMINARY; PRT; 431 AA.
AC Q938V3_STRMU PRELIMINARY; PRT; 431 AA.
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Glucan-binding protein B.

32 65 56.0 544 2 Q840W6_STRMU Q840W6 streptococc
33 65 56.0 544 2 Q840X3_STRMU Q840X3 streptococc
34 64.5 55.6 318 2 Q8G3U4_BIFLO Q8G3U4 bifidobacte
35 64 55.2 215 2 Q4L8G1_STAHJ Q4L8G1 staphylococ
36 64 55.2 249 2 Q4L8P4_STAHJ Q4L8P4 staphylococ
37 62 53.4 59 2 Q5L27_STRT1 Q5L27 streptococc
38 62 53.4 59 2 Q5M2N2_STRT2 Q5M2N2 streptococc
39 61 52.6 157 2 Q5HLU8_STABP Q5HLU8 staphylococ
40 61 52.6 157 2 Q8CMK5_STABP Q8CMK5 staphylococ
41 61 52.6 163 2 Q6GED0_STAAR Q6GED0 staphylococ
42 61 52.6 166 2 Q6G718_STAAS Q6G718 staphylococ
43 61 52.6 166 2 Q5HDQ5_STAAC Q5HDQ5 staphylococ
44 61 52.6 166 2 Q7A060_STAAM Q7A060 staphylococ
45 61 52.6 166 2 Q7A2K7_STAAM Q7A2K7 staphylococ

OS Streptococcus mutans.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1309;

RA NUCLEOTIDE SEQUENCE.
 RC STRAIN=3V4;
 RC Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
 RA "Cloning of the gbpB gene from Streptococcus mutans.";
 RL J. Dent. Res. 79:224-224(2000).
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=3V4;
 RC MEDLINE=21481971; PubMed=11598068;
 RX DOI=10.1128/IAI.69.11.6931-6941.2001;
 RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
 Duncan M.J.;
 RA "Cloning of the Streptococcus mutans gene encoding glucan binding
 RT protein B and analysis of genetic diversity and protein production in
 RT clinical isolates.";
 RL Infect. Immun. 69:6931-6941(2001).
 DR EMBL; AF338445; AAK08104.1; -; Genomic_DNA.
 DR EMBL; AY046410; AAK94500.1; -; Genomic_DNA.
 DR InterPro; IPR007921; CHAP.
 DR InterPro; IPR009148; Siba.
 DR Pfam; PF05257; CHAP; 1.
 DR PRINTS; PR01852; SIBAPROTEIN.
 DR PROSITE; PS50911; CHAP; 1.
 SQ SEQUENCE 431 AA; 44620 MW; 464FE3B563FB7E51 CRC64;

Query Match 89.7%; Score 104; DB 2; Length 431;
 Best Local Similarity 100.0%; Pred. No. 1.6e-05;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNYWNGGQWAAASAAAAG 18
 |||||
 DB 349 GNYWNGGQWAAASAAAAG 366

RESULT 3
 Q9AG98 STRMU PRELIMINARY; PRT; 431 AA.
 AC Q9AG98;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
 DE Immunodominant glycoprotein IDG-60 (Glucan-binding protein B).
 GN Name=saga;
 OS Streptococcus mutans.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1309;
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=GS-5;
 RC MEDLINE=21153617; PubMed=11254612;
 RX DOI=10.1128/IAI.69.4.2493-2501.2001;
 RA Chia J.S., Lee Y.Y., Huang P.T., Chen J.Y.;
 RA "Identification of stress-responsive genes in Streptococcus mutans by
 RT differential display reverse transcription-PCR.";
 RL Infect. Immun. 69:2493-2501(2001).
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=GS-5;
 RC MEDLINE=21481977; PubMed=11598074;
 RX DOI=10.1128/IAI.69.11.6987-6998.2001;
 RA Chia J.S., Chang L.Y., Shun C.T., Chang Y.Y., Chen J.Y.;
 RA "A 60-kilodalton immunodominant glycoprotein is essential for cell
 RT wall integrity and the maintenance of cell shape in Streptococcus
 RT mutans.";
 RL Infect. Immun. 69:6987-6998(2001).
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=GS-5;

RA Chia J.-S., Chang L.-Y., Lee Y.-Y., Chen J.-Y.;
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBSJ databases.
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=3V4;
 RC Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
 RA "Cloning of the gbpB gene from Streptococcus mutans.";
 RL J. Dent. Res. 79:224-224(2000).
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=3V4;
 RC MEDLINE=21481971; PubMed=11598068;
 RX DOI=10.1128/IAI.69.11.6931-6941.2001;
 RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
 Duncan M.J.;
 RA "Cloning of the Streptococcus mutans gene encoding glucan binding
 RT protein B and analysis of genetic diversity and protein production in
 RT clinical isolates.";
 RL Infect. Immun. 69:6931-6941(2001).
 DR EMBL; AF338445; AAK08104.1; -; Genomic_DNA.
 DR EMBL; AY046410; AAK94500.1; -; Genomic_DNA.
 DR InterPro; IPR007921; CHAP.
 DR InterPro; IPR009148; Siba.
 DR Pfam; PF05257; CHAP; 1.
 DR PRINTS; PR01852; SIBAPROTEIN.
 DR PROSITE; PS50911; CHAP; 1.
 SQ SEQUENCE 431 AA; 44592 MW; 3EBE21FC5E47232E CRC64;

Query Match 89.7%; Score 104; DB 2; Length 431;
 Best Local Similarity 100.0%; Pred. No. 1.6e-05;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNYWNGGQWAAASAAAAG 18
 |||||
 DB 349 GNYWNGGQWAAASAAAAG 366

RESULT 4
 Q8DWM3 STRMU PRELIMINARY; PRT; 431 AA.
 AC Q8DWM3;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Putative secreted antigen GbpB/saga; putative peptidoglycan
 DE hydrolase.
 GN Name=gppB; OrderedLocusNames=SMU.22;
 OS Streptococcus mutans.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1309;
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=UA159 / ATCC 700610 / Serotype c;
 RC MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;
 RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.S.P.,
 Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
 Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,
 Ferretti J.J.;
 RA "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
 RT pathogen.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
 DR EMBL; AE014855; AAN57811.1; -; Genomic_DNA.
 DR InterPro; IPR007921; CHAP.
 DR InterPro; IPR009148; Siba.
 DR Pfam; PF05257; CHAP; 1.
 DR PRINTS; PR01852; SIBAPROTEIN.
 DR PROSITE; PS50911; CHAP; 1.
 KW Complete proteome.
 SQ SEQUENCE 431 AA; 44620 MW; 2D1CA695248CCD3E CRC64;

Query Match 89.7%; Score 104; DB 2; Length 431;
 Best Local Similarity 100.0%; Pred. No. 1.6e-05;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNYWNGGQWAASAAAAG 18
|||||
Db 349 GNYWNGGQWAASAAAAG 366

RESULT 5

Q938V1_STRMU
ID Q938V1_STRMU PRELIMINARY; PRT; 432 AA.
AC Q938V1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Glucan-binding protein B.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=3SN1;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gbpB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224(2000).
RN [2]

RN NUCLEOTIDE SEQUENCE.

RC STRAIN=3SN1;
RX MEDLINE=21481971; PubMed=11598068;
RX DOI=10.1128/IAI.69.11.6931-6941.2001;
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
RA Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding
RT protein B and analysis of genetic diversity and protein production in
RT clinical isolates.";
RL Infect. Immun. 69:6931-6941(2001).
DR EMBL; AY046413; AAK94503.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIDA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
SQ SEQUENCE 432 AA; 44652 MW; 3F88ECB9A1F3BE4F CRC64;

Query Match 89.7%; Score 104; DB 2; Length 432;
Best Local Similarity 100.0%; Pred. No. 1.6e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNYWNGGQWAASAAAAG 18
|||||
Db 350 GNYWNGGQWAASAAAAG 367

RESULT 6

Q938V2_STRMU
ID Q938V2_STRMU PRELIMINARY; PRT; 432 AA.
AC Q938V2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Glucan-binding protein B.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=15JP2;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gbpB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224(2000).
RN [2]

RN NUCLEOTIDE SEQUENCE.

RC STRAIN=15JP2;
RX MEDLINE=21481971; PubMed=11598068;
RX DOI=10.1128/IAI.69.11.6931-6941.2001;
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
RA Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding
RT protein B and analysis of genetic diversity and protein production in
RT clinical isolates.";
RL Infect. Immun. 69:6931-6941(2001).
DR EMBL; AY046413; AAK94502.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIDA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
SQ SEQUENCE 432 AA; 44648 MW; E769B2504AEE50E9 CRC64;

Query Match 89.7%; Score 104; DB 2; Length 432;
Best Local Similarity 100.0%; Pred. No. 1.6e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNYWNGGQWAASAAAAG 18
|||||
Db 350 GNYWNGGQWAASAAAAG 367

RESULT 7

Q8E3F4_STRA3
ID Q8E3F4_STRA3 PRELIMINARY; PRT; 169 AA.
AC Q8E3F4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein gba1805.
GN OrderedLocusNames-gba1805;
OS Streptococcus agalactiae (serotype III).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216495;
RN [1]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=NEM316 / Serotype III;
RX MEDLINE=22242508; PubMed=12354221;
RA Glaser P., Rusniok C., Buchrieser C., Chevalier F., Frangeul L.,
RA Msadek T., Zouine M., Couve E., Lalioui L., Poyart C., Trieu-Cuot P.,
RA Kunst F.;
RT "Genome sequence of Streptococcus agalactiae, a pathogen causing
RT invasive neonatal disease.";
RL Mol. Microbiol. 45:1499-1513(2002).
DR EMBL; AL766853; CAD47464.1; -; Genomic_DNA.
DR SAGaList; gba1805; -;
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIDA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 169 AA; 18304 MW; 8A159F753D747869 CRC64;

Query Match 78.4%; Score 91; DB 2; Length 169;
Best Local Similarity 83.3%; Pred. No. 0.00034;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GNYWNGGQWAASAAAAG 18
|||||
Db 79 GNYWNGGQWAASAAAAG 96

RESULT 8

Q8G4C1_BIFLO
ID Q8G4C1_BIFLO PRELIMINARY; PRT; 486 AA.
AC Q8G4C1;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)

```

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Possible TrAg-related protein.
GN OrderedLocusNames=BLI467;
OS Bifidobacterium longum.
OC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
OC Bifidobacteriaceae; Bifidobacterium.
OX NCBI_TaxID=216816;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=NCC 2705;
RX MEDLINE=22294977; PubMed=12381787; DOI=10.1073/pnas.212527599;
RA Schell M.A., Karmirantzou M., Snel B., Vilanova D., Berger B.,
RA Fessi G., Zwahlen M.-C., Desiere F., Bork P., Delley M.,
RA Piedmore R.D., Arigoni F.;
RT "The genome sequence of Bifidobacterium longum reflects its adaptation
RL to the human gastrointestinal tract.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
DR EMBL; AS014295; AAN25262.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR Pfam; PF05257; CHAP; 1.
DR PROSITE; P850911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 486 AA; 51046 MW; D62375650A47B052 CRC64;

Query Match 77.6%; Score 90; DB 2; Length 486;
Best Local Similarity 83.3%; Pred. No. 0.0012;
Matches 15; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GNYWNGGQWASAAAAG 18
DB 230 GSYWNGGDWAASATAAG 247

RESULT 9
Q8DMY4_STR6
ID Q8DMY4_STR6 PRELIMINARY; PRT; 392 AA.
AC Q8DMY4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE General stress protein GSP-781
DE Name=gsp-781; OrderedLocusNames=gsp-781;
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21429245; PubMed=11544234;
RX DOI=10.1128/JB.183.19.5709-5717.2001;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAnnis S.M., McHenry M., McLeaster K., Mundy C.W., Nicas T.I.,
RA Norris F.H., O'Garra M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rosteck P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717(2001).
RL EMBL; AE008566; AAL00823.1; -; Genomic_DNA.
DR PIR; B98124; B98124.
DR PIR; G95258; G95258.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; P850911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 392 AA; 41697 MW; B5A5A860FAEA77DD CRC64;

Query Match 76.7%; Score 89; DB 2; Length 392;
Best Local Similarity 83.3%; Pred. No. 0.0014;
Matches 15; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GNYWNGGQWASAAAAG 18
DB 304 GDYWGNGAQWATSAAAAG 321

RESULT 11
Q8DXT4_STRAS
ID Q8DXT4_STRAS PRELIMINARY; PRT; 169 AA.
AC Q8DXT4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein SAG1762.
DE OrderedLocusNames=SAG1762;
GN Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216466;
RN [1]
RP NUCLEOTIDE SEQUENCE.

```

RC STRAIN=2603 V/R / Serotype V;
 RX MEDLINE=22222988; PubMed=12200547; DOI=10.1073/pnas.182380799;
 RA Tettelein H., Masiagnan V., Cieslewicz M.J., Eisen J.A., Peterson S.N.,
 RA Wessels M.R., Paulsen I.T., Nelson K.E., Margalit I., Read T.D.,
 RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
 RA DeBoy R.T., Durkin A.S., Kolonay J.P., Madupu R., Lewis M.R.,
 RA Radune D., Fedorova N.B., Scanlan D., Khouiri H.M., Mulligan S.,
 RA Carthy H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
 RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Maione D.,
 RA Rinaldo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
 RA Fraser C.M.;
 RT "Complete genome sequence and comparative genomic analysis of an
 RT emerging human pathogen, serotype V *Streptococcus agalactiae*.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).
 DR EMBL; AB014271; ANAN0625.1; -; Genomic_DNA.
 DR TIGR; SAG1762; -;
 DR InterPro; IPR007921; CHAP.
 DR InterPro; IPR009148; Siba.
 DR Pfam; PF05257; CHAP; 1.
 DR PRINTS; PR01852; SIBAPROTEIN.
 DR PROSITE; PS0911; CHAP; 1.
 KW Complete proteome.
 SQ SEQUENCE 169 AA; 18390 MW; 8A1653A8B5B8B769 CRC64;

Query Match 73.3%; Score 85; DB 2; Length 169;
 Best Local Similarity 82.4%; Pred. No. 0.0021;
 Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GNYWNGGQWAAASAAAA 17
 DB 79 GNYWNGGQWAAASAA 95

RESULT 12

OBQRM2 STRT
 ID Q8QRM2_STRT PRELIMINARY; PRT; 125 AA.
 AC Q8QRM2;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein (Fragment).
 OS Streptococcus rattus.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1341;
 RN [1];
 RP STRAIN=BHT;
 RC NUCLEOTIDE SEQUENCE.
 RA Tamura H., Kato H.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB083135; BAB88829.1; -; Genomic DNA.
 DR GO; GO:0016998; P:cell wall catabolism; IEA.
 DR InterPro; IPR007921; CHAP.
 DR InterPro; IPR002482; LysM.
 DR Pfam; PF05257; CHAP; 1.
 DR Pfam; PF01476; LysM; 1.
 DR SMART; SM00257; LysM; 1.
 DR PROSITE; PS0911; CHAP; 1.
 KW Hypothetical protein.
 FT NON TER 125
 FT NON TER 125
 SQ SEQUENCE 125 AA; 12798 MW; A26A0A2C34C1148A CRC64;

Query Match 72.4%; Score 84; DB 2; Length 125;
 Best Local Similarity 77.8%; Pred. No. 0.0021;
 Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 GNYWNGGQWAAASAAAA 18
 DB 89 GNYWNGGQWAAASAA 106

RESULT 13

Q5XEL1 STRP6
 ID Q5XEL1_STRP6 PRELIMINARY; PRT; 398 AA.
 AC Q5XEL1;
 DT 25-OCT-2004 (TrEMBLrel. 28, Created)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE Secreted protein.
 GN OrderedLocusNames=M6_Spy0017;
 OS Streptococcus pyogenes (serotype M6).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=301450;
 RN [1];
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=MGAS10394;
 RX PubMed=15272401; DOI=10.1086/422697;
 RA Banks D.J., Porcella S.F., Barbian K.D., Beres S.B., Phillips L.E.,
 RA Voyich J.M., Deleo F.R., Martin J.M., Somerville G.A., Musher J.M.;
 RT "Progress toward characterization of the group A *Streptococcus*
 RT metagenome: complete genome sequence of a macrolide-resistant serotype
 RT M6 strain.";
 RL J. Infect. Dis. 190:727-738(2004).
 DR EMBL; CP000003; AAT86152.1; -; Genomic_DNA.
 DR InterPro; IPR007921; CHAP.
 DR InterPro; IPR009148; Siba.
 DR Pfam; PF05257; CHAP; 1.
 DR PRINTS; PR01852; SIBAPROTEIN.
 DR PROSITE; PS0911; CHAP; 1.
 KW Complete proteome.
 SQ SEQUENCE 398 AA; 42028 MW; 5DCEDA78CB863B60 CRC64;

Query Match 72.4%; Score 84; DB 2; Length 398;
 Best Local Similarity 83.3%; Pred. No. 0.0064;
 Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GNYWNGGQWAAASAAAA 18
 DB 315 GNYWNGGQWAAASAA 332

RESULT 14
 Q9A1Z8 STRPY
 ID Q9A1Z8_STRPY PRELIMINARY; PRT; 398 AA.
 AC Q9A1Z8; Q7BH59;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
 DE Secreted protein Siba precursor.
 GN OrderedLocusNames=SPY0019;
 OS Streptococcus pyogenes.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1314;
 RN [1];
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=SF370 / ATCC 700294 / Serotype M1;
 RX MEDLINE=21192684; PubMed=11296296; DOI=10.1073/pnas.071559398;
 RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
 RA Primeaux C., Sezate S., Suarov A.N., Kenton S., Lai H.S., Lin S.P.,
 RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
 RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E.;
 RT "Complete genome sequence of an M1 strain of *Streptococcus pyogenes*.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
 RN [2];
 RP NUCLEOTIDE SEQUENCE.
 RA Fagan P.K., Reinscheid D., Gottschalk B., Chhatwal G.S.;
 RT "Identification and characterization of a novel secreted protein from
 RT group A streptococcus.";
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE006474; AAK33158.1; -; Genomic_DNA.
 DR EMBL; AF319999; AAL73135.1; -; Genomic_DNA.
 DR InterPro; IPR007921; CHAP.
 DR InterPro; IPR009148; Siba.


```
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS09111; CHAP; 1.
KW Complete proteome; Signal.
FT SIGNAL 1 23 Potential.
SQ SEQUENCE 398 AA; 41899 MW; 28A9B3F7195E969B CRC64;

Query Match 72.4%; Score 84; DB 2; Length 398;
Best Local Similarity 83.3%; Pred. No. 0.0064;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GNYWNGGOWAASAAAAG 18
DB 315 GNNWNGGOWAYSQAAG 332

RESULT 15
O7CNQ7_STRP8
ID O7CNQ7_STRP8 PRELIMINARY; PRT; 398 AA.
AC O7CNQ7;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Putative secreted protein.
GN Ordered locus names=spyM18_0020;
OS Streptococcus pyogenes (serotype M18).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=301451;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MGAS8232 / Serotype M18;
RX MEDLINE=21927593; PubMed=11917108; DOI=10.1073/pnas.062526099;
RA Smoot J.C., Barbican K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
RA Sylva G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,
RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.;
RT "Genome sequence and comparative microarray analysis of serotype M18
RT group A Streptococcus strains associated with acute rheumatic fever
RT outbreaks.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673 (2002).
DR EMBL; AE009955; AL96849.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS09111; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 398 AA; 42028 MW; 5DCEDA78CB863B60 CRC64;

Query Match 72.4%; Score 84; DB 2; Length 398;
Best Local Similarity 83.3%; Pred. No. 0.0064;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GNYWNGGOWAASAAAAG 18
DB 315 GNNWNGGOWAYSQAAG 332
```

Search completed: February 11, 2006, 19:39:13
Job time : 4.69869 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 18:56:56 ; Search time 2.09028 Seconds
(without alignments)
4414.224 Million cell updates/sec

Title: US-10-797-821-42

Perfect score: 121
Sequence: 1 NNHVSIWEAWSNDTPYLHDD 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001s:*
5: Geneseq2002s:*
6: Geneseq2003as:*
7: Geneseq2003bs:*
8: Geneseq2004s:*
9: Geneseq2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	121	100.0	21	9	Adx37285 Streptoco
2	121	100.0	1590	7	Adp93657 Streptoco
3	121	100.0	1590	9	Adx37280 Streptoco
4	121	100.0	1592	2	Aar32925 Glucosyl
5	115	95.0	21	9	Adv68564 S. sobrin
6	115	95.0	21	9	Adv68562 S. downi
7	110	90.9	1017	5	Aau79285 Streptoco
8	110	90.9	1475	5	Aau98031 S. mutans
9	110	90.9	1475	5	Aau98036 S. mutans
10	110	90.9	1475	5	Aau98037 S. mutans
11	110	90.9	1475	5	Aau98040 S. mutans
12	110	90.9	1475	5	Aau98035 S. mutans
13	110	90.9	1475	5	Aau98033 S. mutans
14	110	90.9	1475	5	Aau98034 S. mutans
15	110	90.9	1475	5	Aau98030 S. mutans
16	110	90.9	1475	5	Aau98039 S. mutans
17	110	90.9	1475	5	Aau98027 S. mutans
18	110	90.9	1475	5	Aau98032 S. mutans
19	110	90.9	1475	5	Aau98038 S. mutans
20	110	90.9	1475	7	Adp93654 Streptoco
21	110	90.9	1475	9	Adx37277 Streptoco
22	110	90.9	1476	5	Aau79284 Streptoco
23	104	86.0	21	9	Adv68551 S. mutans
24	104	86.0	21	9	Adv68560 S. mutans

25	101	83.5	1375	5	Aau98028 S. mutans
26	101	83.5	1375	5	Aau79288 Streptoco
27	101	83.5	1375	7	Adp93655 Streptoco
28	101	83.5	1375	9	Adx37278 Streptoco
29	76	62.8	1499	7	Adc54806 Protein S
30	73	60.3	21	9	Adv68561 S. mutans
31	73	60.3	1430	5	Aau98043 S. mutans
32	73	60.3	1430	5	Aau98044 S. mutans
33	73	60.3	1430	5	Aau98045 S. mutans
34	73	60.3	1430	5	Aau98042 S. mutans
35	73	60.3	1430	5	Aau98029 S. mutans
36	73	60.3	1430	5	Aau98041 S. mutans
37	73	60.3	1430	7	Adp93656 Streptoco
38	73	60.3	1430	9	Adx37279 Streptoco
39	68	56.2	223	6	Abg63230 Glucansuc
40	68	56.2	2835	5	Abb98574 Dextran s
41	68	56.2	2835	6	Abb55594 Amino aci
42	66	54.5	12	5	Abb98642 Dextrane-
43	66	54.5	1365	7	Adp93659 Streptoco
44	66	54.5	1365	9	Adx37282 Streptoco
45	64	52.9	223	6	Abg63228 Glucansuc

ALIGNMENTS

RESULT 1
ADx37285
ID ADx37285 standard; peptide; 21 AA.
XX
AC ADx37285;
XX
DT 21-APR-2005 (first entry)
XX
DE Streptococcus sobrinus glucan binding protein B peptide #1.
XX
KW immunogenicity; immune stimulation; glucan binding protein-B;
KW microparticle; major histocompatibility complex; tooth disease.
XX
OS Streptococcus sobrinus.
XX
PN US2005031633-A1.
XX
PD 10-FEB-2005.
XX
PF 09-MAR-2004; 2004US-00797821.
XX
PR 13-APR-1998; 98US-0081550P.
PR 08-JAN-1999; 99US-0115142P.
PR 12-APR-1999; 99US-00290049.
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
PR 07-MAR-2003; 2003US-00383930.
(SMIT/) SMITH D J.
(TAUB/) TAUBMAN M A.
XX
PI Smith DJ, Taubman MA;
XX
WPI; 2005-151644/16.
XX
New composition comprising a fragment of a glucan binding protein-B (GbpB) that binds to MHC class II protein, and a biocompatible microparticle, useful for producing an antibody (claimed) for immunizing mammals against dental caries.
XX
Example 5; SEQ ID NO 42; 73pp; English.
XX
The invention relates to a composition comprising a fragment of a glucan binding protein-B (GbpB) and a biocompatible microparticle, where the fragment binds to a major histocompatibility complex (MHC) class II protein. The composition is useful for producing an antibody for immunizing mammals against dental caries. This sequence corresponds to a

```

CC Streptococcus sobrinus GbpB-derived peptide of the invention.
XX
SQ Sequence 21 AA;

  Query Match      100.0%; Score 121; DB 9; Length 21;
  Best Local Similarity 100.0%; Pred. No. 7, 9e-11;
  Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNHVSIVEAWSNDTPYLHDD 21
   |||||
Db 1 NNHVSIVEAWSNDTPYLHDD 21

RESULT 2
ADD93657
ID ADD93657 standard; protein; 1590 AA.
XX
AC ADD93657;
XX
DT 29-JAN-2004 (first entry)
XX
DE Streptococcus sobrinus glucosyltransferase-I.
XX
KW Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.
XX
OS Streptococcus sobrinus.
XX
PN WQ2003075845-A2.
XX
PD 18-SEP-2003.
XX
PF 07-MAR-2003; 2003WO-US006962.
XX
PR 07-MAR-2002; 2002US-0363209P.
XX
PR 08-AUG-2002; 2002US-0402483P.
XX
PA (FORS-) FORSYTH INST.
XX
PI Smith DJ, Taubman MA;
XX
PI WPI; 2003-845091/78.
XX
DR Composition useful as vaccines for dental caries comprises a fragment of
PT a glucan binding protein-B binding to a major histocompatibility complex
PT class II protein.
XX
PS Claim 16; Page 14; 49pp; English.
XX
CC The present sequence is the protein sequence of Streptococcus sobrinus
CC glucosyltransferase-I (GTF-I). Peptide fragments of GTF-I, especially
CC from the catalytic domain of the polypeptide, can be used in immunogenic
CC compositions and subunit vaccines for dental caries. These compositions
CC comprise a major histocompatibility complex (MHC) class II protein-
CC binding peptide from S. mutants glucan binding protein-B (GbpB)
CC covalently linked with a peptide fragment of a streptococcal
CC glucosyltransferase. The compositions are used in a claimed method of
CC eliciting production of an antibody in a mammal. Dieptopic or
CC multiepitopic polypeptides can be prepared synthetically or by
CC recombinant DNA technology. Antibodies raised against MHC class II
CC binding fragments of GbpB can be used in passive immunisation.
XX
SQ Sequence 1590 AA;

  Query Match      100.0%; Score 121; DB 7; Length 1590;
  Best Local Similarity 100.0%; Pred. No. 1,1e-08;
  Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNHVSIVEAWSNDTPYLHDD 21
   |||||
Db 478 NNHVSIVEAWSNDTPYLHDD 498

RESULT 3
ADD93657
ID ADD93657 standard; protein; 1590 AA.
XX
AC ADD93657;
XX
DT 29-JAN-2004 (first entry)
XX
DE Streptococcus sobrinus glucosyltransferase-I.
XX
KW Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.
XX
OS Streptococcus sobrinus.
XX
PN WQ2003075845-A2.
XX
PD 18-SEP-2003.
XX
PF 07-MAR-2003; 2003WO-US006962.
XX
PR 07-MAR-2002; 2002US-0363209P.
XX
PR 08-AUG-2002; 2002US-0402483P.
XX
PA (FORS-) FORSYTH INST.
XX
PI Smith DJ, Taubman MA;
XX
PI WPI; 2003-845091/78.
XX
DR Composition useful as vaccines for dental caries comprises a fragment of
PT a glucan binding protein-B binding to a major histocompatibility complex
PT class II protein.
XX
PS Claim 16; Page 14; 49pp; English.
XX
CC The present sequence is the protein sequence of Streptococcus sobrinus
CC glucosyltransferase-I (GTF-I). Peptide fragments of GTF-I, especially
CC from the catalytic domain of the polypeptide, can be used in immunogenic
CC compositions and subunit vaccines for dental caries. These compositions
CC comprise a major histocompatibility complex (MHC) class II protein-
CC binding peptide from S. mutants glucan binding protein-B (GbpB)
CC covalently linked with a peptide fragment of a streptococcal
CC glucosyltransferase. The compositions are used in a claimed method of
CC eliciting production of an antibody in a mammal. Dieptopic or
CC multiepitopic polypeptides can be prepared synthetically or by
CC recombinant DNA technology. Antibodies raised against MHC class II
CC binding fragments of GbpB can be used in passive immunisation.
XX
SQ Sequence 1590 AA;

  Query Match      100.0%; Score 121; DB 9; Length 1590;
  Best Local Similarity 100.0%; Pred. No. 1,1e-08;
  Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNHVSIVEAWSNDTPYLHDD 21
   |||||
Db 478 NNHVSIVEAWSNDTPYLHDD 498

RESULT 4
AAR32925
ID AAR32925 standard; protein; 1592 AA.
XX
AC AAR32925;
XX
DT 28-JUN-1993 (first entry)
XX
DE Glucosyltransferase I.
XX
KW GT-1; Streptococcus; dental; caries.
XX
OS Streptococcus sobrinus.
XX

```

PN JP05023188-A.
 XX
 PD 02-FEB-1993.
 XX
 PF 25-JUL-1991; | 91JP-00186592.
 XX
 PR 25-JUL-1991; 91JP-00186592.
 XX
 XX (KATO/) KATO K.
 PA (FUKU/) FUKUI I.
 XX
 DR WPI; 1993-079449/10.
 DR N-PSDB; AAQ37760.
 XX
 FT DNA sequence glucosyltransferase-I - comprises Streptococcus sobrinus
 PT DNA sequence with at least one nucleotide added or deleted.
 XX
 XX Claim 13; Page 15; 29pp; Japanese.
 XX
 CC The DNA sequence from Streptococcus sobrinus strain 6715 encodes
 CC glucosyltransferase-I (and mutants). The DNA was obtd. by treating S.
 CC sobrinus 6715 with mutanolysin, extracting the chromosomal DNA, partially
 CC digesting with Sau3AI and fractionating on agarose gel. The 3-5 kbp
 CC fragment was ligated into pUC18 and E. coli JM109 transformed with it. A
 CC GT-1 expressing clone was isolated and sequenced. The clone may be used
 CC in the development of a drug for dental caries
 XX
 SQ Sequence 1592 AA;
 Query Match 100.0%; Score 121; DB 2; Length 1592;
 Best Local Similarity 100.0%; Pred. No. 1.1e-08;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NNHVSIVEAWSNDTPYLHDD 21
 |||||
 DB 478 NNHVSIVEAWSNDTPYLHDD 498
 RESULT 5
 ADV68564
 ID ADV68564 standard; peptide; 21 AA.
 XX
 AC ADV68564; |
 XX
 DT 24-FEB-2005 (first entry)
 XX
 DE S. sobrinus GTF2 EAW peptide.
 XX
 KW Immunogenicity; vaccine; glucosyltransferase; GTF; tooth disease;
 KW mouth disease; antibacterial; EAW peptide.
 XX
 OS Streptococcus sobrinus.
 XX
 PN US6827936-B1.
 XX
 PD 07-DEC-2004.
 XX
 PF 12-APR-1999; 99US-00290049.
 XX
 PR 13-APR-1998; 98US-0081550P.
 PR 08-JAN-1999; 99US-0115142P.
 XX
 PA (FORS-) FORSYTH DENTAL INFIRMARY FOR CHILDREN.
 XX
 PI Smith DJ, Taubman MA;
 XX
 DR WPI; 2005-009976/01.
 XX
 PT New immunogenic composition comprises a peptide corresponding to a
 PT subunit of glucosyltransferase, useful as glucosyltransferase subunit
 PT vaccine for preventing dental caries.
 XX
 PS Claim 2; SEQ ID NO 14; 18pp; English.

XX The present invention relates to a novel immunogenic composition which
 CC comprises a peptide corresponding to a subunit of glucosyltransferase
 CC (GTF) enzyme. The immunogenic composition is useful as a GTF subunit
 CC vaccine for preventing dental caries. The present sequence is the
 CC Streptococcus sobrinus GTF2 EAW [(beta, alpha)8 barrel domain beta 5
 CC strand associated catalytic region] peptide.
 XX
 SQ Sequence 21 AA;
 Query Match 95.0%; Score 115; DB 9; Length 21;
 Best Local Similarity 100.0%; Pred. No. 6.1e-10;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NNHVSIVEAWSNDTPYLHD 20
 |||||
 DB 2 NNHVSIVEAWSNDTPYLHD 21
 RESULT 6
 ADV68562
 ID ADV68562 standard; peptide; 21 AA.
 XX
 AC ADV68562;
 XX
 DT 24-FEB-2005 (first entry)
 XX
 DE S. downei GTF-I EAW peptide.
 XX
 KW Immunogenicity; vaccine; glucosyltransferase; GTF; tooth disease;
 KW mouth disease; antibacterial; EAW peptide.
 XX
 OS Streptococcus downei.
 XX
 PN US6827936-B1.
 XX
 PD 07-DEC-2004.
 XX
 PF 12-APR-1999; 99US-00290049.
 XX
 PR 13-APR-1998; 98US-0081550P.
 PR 08-JAN-1999; 99US-0115142P.
 XX
 PA (FORS-) FORSYTH DENTAL INFIRMARY FOR CHILDREN.
 XX
 PI Smith DJ, Taubman MA;
 XX
 DR WPI; 2005-009976/01.
 XX
 PT New immunogenic composition comprises a peptide corresponding to a
 PT subunit of glucosyltransferase, useful as glucosyltransferase subunit
 PT vaccine for preventing dental caries.
 XX
 PS Claim 2; SEQ ID NO 12; 18pp; English.
 XX
 CC The present invention relates to a novel immunogenic composition which
 CC comprises a peptide corresponding to a subunit of glucosyltransferase
 CC (GTF) enzyme. The immunogenic composition is useful as a GTF subunit
 CC vaccine for preventing dental caries. The present sequence is the
 CC Streptococcus downei GTF-I EAW [(beta, alpha)8 barrel domain beta 5 strand
 CC associated catalytic region] peptide.
 XX
 SQ Sequence 21 AA;
 Query Match 95.0%; Score 115; DB 9; Length 21;
 Best Local Similarity 100.0%; Pred. No. 6.1e-10;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NNHVSIVEAWSNDTPYLHD 20
 |||||
 DB 2 NNHVSIVEAWSNDTPYLHD 21

Qy 1 NNHVSIVEAWSNDTPYLHDD 21
 Db 482 NDHLSILEAWSNDTPYLHDD 502

RESULT 9
 AAU98036
 ID AAU98036 standard; protein; 1475 AA.
 AC AAU98036;
 XX
 DT 27-AUG-2002 (first entry)
 XX
 DE S. mutans glucosyltransferase GTFB mutant D567T/D571K.
 XX
 KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;
 KW coating composition; glucan; starch; latex; thermoplastic molecule;
 KW amyloplast; vacuole; paper manufacture; mutant; mutein.
 XX
 OS Streptococcus mutans.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 567 /note= "Wild-type Asp substituted by Thr"
 FT Misc-difference 571 /note= "Wild-type Asp substituted by Lys"
 FT US2002031826-A1.
 PN 14-MAR-2002.
 PD
 PF 19-DEC-2000; 2000US-00740274.
 PR 07-JUN-1995; 95US-00478704.
 PR 07-JUN-1995; 95US-00482711.
 PR 16-JAN-1998; 98US-00007999.
 PR 16-JAN-1998; 98US-00008172.
 PR 20-JAN-1998; 98US-00009620.
 PR 11-DEC-1998; 98US-00210361.
 XX
 PA (NICH/) NICHOLS S E.
 XX Nichols SE;
 XX WPI; 2002-414332/44.
 XX
 FT Glucosyltransferase B or D protein useful for producing a glucan useful
 FT as substitutes for and additions to modified starch and latexes in paper
 FT manufacture, comprises mutations in specific positions.
 XX
 PS Claim 36; Page; 4app; English.
 XX
 CC The invention an isolated protein comprising a glucosyltransferase (GTF)
 CC B polypeptide having changes at position from I48V, D457N, D567T,
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/K1014T,
 CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the
 CC GTF mutant, an expression cassette comprising the polynucleotide operably
 CC linked to a promoter, a vector comprising the expression cassette, host
 CC cell introduced with the vector, a transgenic plant comprising the
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
 CC coating composition comprising a glucan produced in a plant transformed
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
 CC thermoplastic molecule or their combinations or glucan and starch where
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line
 CC deficient in starch biosynthesis, transformed with a gene encoding a
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper

comprising the glucan (paper sizing/coating agent). The vector is useful
 for producing a glucan in a plant. The method comprises transforming a
 plant cell with the vector, growing the plant cell under plant growing
 conditions to produce a regenerated plant and inducing expression of the
 polynucleotide for a time sufficient to produce the glucan in the
 regenerated plant, where the vector contains a transit sequence from
 ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
 chlorophyll AB binding protein to produce a transgenic plant, and glucan
 is produced in the amyloplast of potato or the vacuole of sugar beet.
 Glucans are useful as substitutes for and additions to modified starch
 and latexes in paper manufacture. Unlike prior art techniques, which
 require input materials that produce chemical effluents, paper
 manufacture utilising the glucan produced by GTF, which utilises
 biologically produced input materials, is more cost-effective and
 environmentally friendly. Moreover, glucans also exhibit thermoplastic
 properties and impart gloss to the paper during coating step. The present
 sequence represents a GTFB mutant of the invention. Note: The present
 sequence is not shown in the specification but was created by the indexer
 using the GTFB sequence appearing as AAU98027 and the information in
 claim 36

XX Sequence 1475 AA;
 SQ
 Query Match 90.9%; Score 110; DB 5; Length 1475;
 Best Local Similarity 85.7%; Pred. No. 4.5e-07;
 Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 NNHVSIVEAWSNDTPYLHDD 21
 Db 482 NDHLSILEAWSNDTPYLHDD 502

RESULT 10
 AAU98037
 ID AAU98037 standard; protein; 1475 AA.
 AC AAU98037;
 XX
 DT 27-AUG-2002 (first entry)
 XX
 DE S. mutans glucosyltransferase GTFB mutant D567T/D571K/K1014T.
 XX
 KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;
 KW coating composition; glucan; starch; latex; thermoplastic molecule;
 KW amyloplast; vacuole; paper manufacture; mutant; mutein.
 XX
 OS Streptococcus mutans.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 567 /note= "Wild-type Asp substituted by Thr"
 FT Misc-difference 571 /note= "Wild-type Asp substituted by Lys"
 FT Misc-difference 1014 /note= "Wild-type Lys substituted by Thr"
 FT US2002031826-A1.
 PN 14-MAR-2002.
 PD
 PF 19-DEC-2000; 2000US-00740274.
 PR 07-JUN-1995; 95US-00478704.
 PR 07-JUN-1995; 95US-00482711.
 PR 16-JAN-1998; 98US-00007999.
 PR 16-JAN-1998; 98US-00008172.
 PR 20-JAN-1998; 98US-00009620.
 PR 11-DEC-1998; 98US-00210361.
 XX
 PA (NICH/) NICHOLS S E.
 XX Nichols SE;
 XX WPI; 2002-414332/44.
 XX
 FT Glucosyltransferase B or D protein useful for producing a glucan useful
 FT as substitutes for and additions to modified starch and latexes in paper
 FT manufacture, comprises mutations in specific positions.
 XX
 PS Claim 36; Page; 4app; English.
 XX
 CC The invention an isolated protein comprising a glucosyltransferase (GTF)
 CC B polypeptide having changes at position from I48V, D457N, D567T,
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/K1014T,
 CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the
 CC GTF mutant, an expression cassette comprising the polynucleotide operably
 CC linked to a promoter, a vector comprising the expression cassette, host
 CC cell introduced with the vector, a transgenic plant comprising the
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
 CC coating composition comprising a glucan produced in a plant transformed
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
 CC thermoplastic molecule or their combinations or glucan and starch where
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line
 CC deficient in starch biosynthesis, transformed with a gene encoding a
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper

PI Nichols SE;
 XX WPI; 2002-414332/44.
 DR
 XX Glucosyltransferase B or D protein useful for producing a glucan useful
 PT as substitutes for and additions to modified starch and latexes in paper
 PT manufacture, comprises mutations in specific positions.
 PT
 XX
 PS Claim 36; Page; 44pp; English.
 XX
 CC The invention an isolated protein comprising a glucosyltransferase (GTF)
 CC B polypeptide having changes at position from 1448V, D457N, D567T,
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,
 CC I448V/D457N/D567T/D571K/K799Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the
 CC GTF mutant, an expression cassette comprising the polynucleotide operably
 CC linked to a promoter, a vector comprising the expression cassette, host
 CC cell introduced with the vector, a transgenic plant comprising the
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
 CC coating composition comprising a glucan produced in a plant transformed
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
 CC thermoplastic molecule or their combinations or glucan and starch where
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line
 CC deficient in starch biosynthesis, transformed with a gene encoding a
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
 CC comprising the glucan (paper sizing/coating agent). The vector is useful
 CC for producing a glucan in a plant. The method comprises transforming a
 CC plant cell with the vector, growing the plant cell under plant growing
 CC conditions to produce a regenerated plant and inducing expression of the
 CC polynucleotide for a time sufficient to produce the glucan in the
 CC regenerated plant, where the vector contains a transit sequence from
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.
 CC Glucans are useful as substitutes for and additions to modified starch
 CC and latexes in paper manufacture. Unlike prior art techniques, which
 CC require input materials that produce chemical effluents, paper
 CC manufacture utilising the glucan produced by GTF, which utilises
 CC biologically produced input materials, is more cost-effective and
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
 CC properties and impart gloss to the paper during coating step. The present
 CC sequence represents a GTFB mutant of the invention. Note: The present
 CC sequence is not shown in the specification but was created by the indexer
 CC using the GTFB sequence appearing as AAU98027 and the information in
 CC claim 36
 XX
 SQ Sequence 1475 AA;
 Query Match 90.9%; Score 110; DB 5; Length 1475;
 Best Local Similarity 85.7%; Pred. No. 4.5e-07;
 Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NNHVSIVEAWSNDPTPLHDD 21
 Db 482 NDHLSILEAWSNDPTPLHDD 502
 RESULT 11
 AAU98040
 ID AAU98040 standard; protein; 1475 AA.
 XX
 AC AAU98040;
 XX
 DT 27-AUG-2002 (first entry)
 XX
 DE S. mutans glucosyltransferase GTFB mutant K779Q.
 XX
 KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;
 KW coating composition; glucan; starch; latex; thermoplastic molecule;
 KW amyloplast; vacuole; paper manufacture; mutant; muitein.

XI Streptococcus mutans.
 OS Synthetic.
 XX
 XX Key Location/Qualifiers
 FT Misc-difference 779
 FT /note= "Wild-type Lys substituted by Gln"
 XX
 PN US2002031826-A1.
 XX
 XX 14-MAR-2002.
 XX
 PF 19-DEC-2000; 2000US-00740274.
 XX
 PR 07-JUN-1995; 95US-00478704.
 PR 07-JUN-1995; 95US-00482711.
 PR 07-JUN-1995; 95US-00485243.
 PR 16-JAN-1998; 98US-00007999.
 PR 16-JAN-1998; 98US-00008172.
 PR 20-JAN-1998; 98US-00009620.
 PR 11-DEC-1998; 98US-00210361.
 XX
 PA (NICH/) NICHOLS S E.
 XX
 PI Nichols SE;
 XX
 XX WPI; 2002-414332/44.
 DR
 XX Glucosyltransferase B or D protein useful for producing a glucan useful
 PT as substitutes for and additions to modified starch and latexes in paper
 PT manufacture, comprises mutations in specific positions.
 PT
 XX
 XX Claim 36; Page; 44pp; English.
 XX
 CC The invention an isolated protein comprising a glucosyltransferase (GTF)
 CC B polypeptide having changes at position from 1448V, D457N, D567T,
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,
 CC I448V/D457N/D567T/D571K/K799Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the
 CC GTF mutant, an expression cassette comprising the polynucleotide operably
 CC linked to a promoter, a vector comprising the expression cassette, host
 CC cell introduced with the vector, a transgenic plant comprising the
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
 CC coating composition comprising a glucan produced in a plant transformed
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
 CC thermoplastic molecule or their combinations or glucan and starch where
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line
 CC deficient in starch biosynthesis, transformed with a gene encoding a
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
 CC comprising the glucan (paper sizing/coating agent). The vector is useful
 CC for producing a glucan in a plant. The method comprises transforming a
 CC plant cell with the vector, growing the plant cell under plant growing
 CC conditions to produce a regenerated plant and inducing expression of the
 CC polynucleotide for a time sufficient to produce the glucan in the
 CC regenerated plant, where the vector contains a transit sequence from
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.
 CC Glucans are useful as substitutes for and additions to modified starch
 CC and latexes in paper manufacture. Unlike prior art techniques, which
 CC require input materials that produce chemical effluents, paper
 CC manufacture utilising the glucan produced by GTF, which utilises
 CC biologically produced input materials, is more cost-effective and
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
 CC properties and impart gloss to the paper during coating step. The present
 CC sequence represents a GTFB mutant of the invention. Note: The present
 CC sequence is not shown in the specification but was created by the indexer
 CC using the GTFB sequence appearing as AAU98027 and the information in
 CC claim 36
 XX
 SQ Sequence 1475 AA;

SQ Sequence 1475 AA;

Query Match 90.9%; Score 110; DB 5; Length 1475;
 Best Local Similarity 85.7%; Pred. No. 4.5e-07;
 Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNHVSIVEAWSNDNTPYLHDD 21
 Db 482 NDHLSILEAWSNDNTPYLHDD 502

RESULT 12
 AAU98035
 ID AAU98035 standard; protein; 1475 AA.
 XX
 AC AAU98035;
 XX
 DT 27-AUG-2002 (first entry)
 XX
 DE S. mutans glucosyltransferase GTFB mutant D457N/D571K.
 XX
 KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;
 KW coating composition; glucan; starch; latex; thermoplastic molecule;
 KW amyloplast; vacuole; paper manufacture; mutant; mutein.
 XX
 OS Streptococcus mutans.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 457 /note= "Wild-type Asp substituted by Asn"
 FT Misc-difference 571 /note= "Wild-type Asp substituted by Lys"
 FT XX
 PN US2002031826-A1.
 XX
 PD 14-MAR-2002.
 XX
 PF 19-DEC-2000; 2000US-00740274.
 XX
 PR 07-JUN-1995; 95US-00478704.
 PR 07-JUN-1995; 95US-00482711.
 PR 07-JUN-1995; 95US-00485243.
 PR 16-JAN-1998; 98US-00007999.
 PR 16-JAN-1998; 98US-00008172.
 PR 20-JAN-1998; 98US-00009620.
 PR 11-DEC-1998; 98US-00210361.
 XX
 PA (NICH/) NICHOLS S E.
 XX
 PI Nichols SE;
 XX
 DR WPI; 2002-414332/44.
 XX
 PT Glucosyltransferase B or D protein useful for producing a glucan useful
 PT as substitutes for and additions to modified starch and latexes in paper
 PT manufacture, comprises mutations in specific positions.
 XX
 PS Claim 36; Page; 44pp; English.
 XX

The invention an isolated protein comprising a glucosyltransferase (GTF)
 B polypeptide having changes at position from I448V, D457N, D567T,
 K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/K1014T,
 I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
 GTF D polypeptide having changes at positions from T589D, T589E, N471D,
 N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
 GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
 complementary polynucleotide, a ribonucleic acid sequence encoding the
 GTF mutant, an expression cassette comprising the polynucleotide operably
 linked to a promoter, a vector comprising the expression cassette, host
 cell introduced with the vector, a transgenic plant comprising the
 vector, a seed or tuber from the transgenic plant, a paper sizing and/or
 coating composition comprising a glucan produced in a plant transformed

CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
 CC thermoplastic molecule or their combinations or glucan and starch where
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line
 CC deficient in starch biosynthesis, transformed with a gene encoding a
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
 CC comprising the glucan (paper sizing/coating agent). The vector is useful
 CC for producing a glucan in a plant. The method comprises transforming a
 CC plant cell with the vector, growing the plant cell under plant growing
 CC conditions to produce a regenerated plant and inducing expression of the
 CC polynucleotide for a time sufficient to produce the glucan in the
 CC regenerated plant, where the vector contains a transit sequence from
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.
 CC Glucans are useful as substitutes for and additions to modified starch
 CC and latexes in paper manufacture. Unlike prior art techniques, which
 CC require input materials that produce chemical effluents, paper
 CC manufacture utilising the glucan produced by GTF, which utilises
 CC biologically produced input materials, is more cost-effective and
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
 CC properties and impart gloss to the paper during coating step. The present
 CC sequence represents a GTFB mutant of the invention. Note: The present
 CC sequence is not shown in the specification but was created by the indexer
 CC using the GTFB sequence appearing as AAU98027 and the information in
 CC claim 36
 CC
 XX
 SQ Sequence 1475 AA;

Query Match 90.9%; Score 110; DB 5; Length 1475;
 Best Local Similarity 85.7%; Pred. No. 4.5e-07;
 Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNHVSIVEAWSNDNTPYLHDD 21
 Db 482 NDHLSILEAWSNDNTPYLHDD 502

RESULT 13
 AAU98033
 ID AAU98033 standard; protein; 1475 AA.
 XX
 AC AAU98033;
 XX
 DT 27-AUG-2002 (first entry)
 XX
 DE S. mutans glucosyltransferase GTFB mutant K1014T.
 XX
 KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;
 KW coating composition; glucan; starch; latex; thermoplastic molecule;
 KW amyloplast; vacuole; paper manufacture; mutant; mutein.
 XX
 OS Streptococcus mutans.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 1014 /note= "Wild-type Lys substituted by Thr"
 FT XX
 PN US2002031826-A1.
 XX
 PD 14-MAR-2002.
 XX
 PF 19-DEC-2000; 2000US-00740274.
 XX
 PR 07-JUN-1995; 95US-00478704.
 PR 07-JUN-1995; 95US-00482711.
 PR 07-JUN-1995; 95US-00485243.
 PR 16-JAN-1998; 98US-00007999.
 PR 16-JAN-1998; 98US-00008172.
 PR 20-JAN-1998; 98US-00009620.
 PR 11-DEC-1998; 98US-00210361.
 XX
 PA (NICH/) NICHOLS S E.

CC using the GTFB sequence appearing as AAU98027 and the information in
 CC claim 36

SQ Sequence 1475 AA;

Query Match 90.9%; Score 110; DB 5; Length 1475;

Best Local Similarity 85.7%; Pred. No. 4.5e-07; Indels 0; Gaps 0;
 Matches 18; Conservative 3; Mismatches 0;

QY 1 NNHVSTVEAWSNDTPYLHDD 21

Db 482 NDHLSLEAWSNDTPYLHDD 502

RESULT 15

AAU98030

ID AAU98030 standard; protein; 1475 AA.

AC AAU98030;

DT 27-AUG-2002 (first entry)

DE S. mutans glucosyltransferase GTFB mutant I448V.

KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;

XW coating composition; glucan; starch; latex; thermoplastic molecule;

XX amyloplast; vacuole; paper manufacture; mutant; mutein.

OS Streptococcus mutans.

XX Synthetic.

FH Key Location/Qualifiers

FT Misc-difference 448

FT /note= "Wild-type Ile substituted by Val"

XX US2002031826-A1.

XX PD 14-MAR-2002.

XX PF 19-DEC-2000; 2000US-00740274.

XX PR 07-JUN-1995; 95US-00478704.

XX PR 07-JUN-1995; 95US-00482711.

XX PR 07-JUN-1995; 95US-00485243.

XX PR 16-JAN-1998; 98US-00007999.

XX PR 16-JAN-1998; 98US-00008172.

XX PR 20-JAN-1998; 98US-00009620.

XX PR 11-DEC-1998; 98US-00210361.

XX (NICH/) NICHOLS S E.

XX Nichols SE;

XX WPI; 2002-414332/44.

XX Glucosyltransferase B or D protein useful for producing a glucan useful
 PT as substitutes for and additions to modified starch and latexes in paper
 PT manufacture, comprises mutations in specific positions.

PS Claim 36; Page; 44pp; English.

XX The invention an isolated protein comprising a glucosyltransferase (GTF)
 CC B polypeptide having changes at position from I448V, D457N, D567T,
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,
 CC I448V/D457N/D567T/D571K/K79Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the
 CC GTF mutant, an expression cassette comprising the polynucleotide operably
 CC linked to a promoter, a vector comprising the expression cassette, host
 CC cell introduced with the vector, a transgenic plant comprising the
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or

CC coating composition comprising a glucan produced in a plant transformed
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
 CC thermoplastic molecule or their combinations or glucan and starch where
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line
 CC deficient in starch biosynthesis, transformed with a gene encoding a
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
 CC comprising the glucan (paper sizing/coating agent). The vector is useful
 CC for producing a glucan in a plant. The method comprises transforming a
 CC plant cell with the vector, growing the plant cell under plant growing
 CC conditions to produce a regenerated plant and inducing expression of the
 CC polynucleotide for a time sufficient to produce the glucan in the
 CC regenerated plant, where the vector contains a transit sequence from
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.
 CC Glucans are useful as substitutes for and additions to modified starch
 CC and latexes in paper manufacture. Unlike prior art techniques, which
 CC require input materials that produce chemical effluents, paper
 CC manufacture utilising the glucan produced by GTF, which utilises
 CC biologically produced input materials, is more cost-effective and
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
 CC properties and impart gloss to the paper during coating step. The present
 CC sequence represents a GTFB mutant of the invention. Note: The present
 CC sequence is not shown in the specification but was created by the indexer
 CC using the GTFB sequence appearing as AAU98027 and the information in
 CC claim 36

XX SQ Sequence 1475 AA;

Query Match 90.9%; Score 110; DB 5; Length 1475;

Best Local Similarity 85.7%; Pred. No. 4.5e-07;

Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNHVSTVEAWSNDTPYLHDD 21

Db 482 NDHLSLEAWSNDTPYLHDD 502

Search completed: February 11, 2006, 19:15:28

Job time : 2.09028 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 19:39:30 ; Search time 0.45978 Seconds
(without alignments)
3776.130 Million cell updates/sec

Title: US-10-797-821-42

Perfect score: 121

Sequence: 1 NNHVSIVEAWSNDTPYLHDD 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/1/iaa/5 COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/6 COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/H COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/PCTUS COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/RE COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	115	95.0	21	2	US-09-290-049A-12
2	115	95.0	21	2	US-09-290-049A-14
3	110	90.9	1475	2	US-09-007-999-2
4	110	90.9	1475	2	US-09-210-361-2
5	110	90.9	1475	2	US-09-740-274-2
6	104	86.0	21	2	US-09-290-049A-1
7	104	86.0	21	2	US-09-290-049A-10
8	101	83.5	1375	2	US-09-210-361-4
9	101	83.5	1375	2	US-09-740-274-4
10	73	60.3	21	2	US-09-290-049A-11
11	73	60.3	545	2	US-09-604-957-4
12	73	60.3	545	2	US-09-995-749A-10
13	73	60.3	1430	2	US-09-008-172-2
14	73	60.3	1430	2	US-09-210-361-6
15	73	60.3	1430	2	US-09-740-274-6
16	61	50.4	21	2	US-09-290-049A-13
17	61	50.4	522	2	US-09-995-749A-11
18	61	50.4	523	2	US-09-604-957-5
19	57	47.1	1577	1	US-08-793-824-2
20	51	42.1	584	2	US-09-604-957-6
21	51	42.1	584	2	US-09-995-749A-12
22	51	42.1	2057	2	US-09-499-203-2
23	47	38.8	535	2	US-09-604-957-7
24	47	38.8	535	2	US-09-995-749A-13
25	47	38.8	749	2	US-09-562-737-96
26	47	38.8	1278	2	US-09-604-957-3
27	47	38.8	1781	2	US-09-995-749A-2

28	47	38.8	2465	1	US-08-596-291-3	Sequence 3, Appli
29	47	38.8	2465	2	US-09-100-804-3	Sequence 3, Appli
30	47	38.8	2466	2	US-09-080-855-12	Sequence 12, Appli
31	47	38.8	2466	2	US-09-566-076-12	Sequence 12, Appli
32	47	38.8	2466	4	PCT-US94-09943-2	Sequence 2, Appli
33	47	38.8	2485	2	US-09-290-640-46	Sequence 46, Appli
34	47	38.8	2485	2	US-09-665-615B-46	Sequence 46, Appli
35	46	38.0	261	2	US-09-902-540-13119	Sequence 13119, A
36	46	38.0	574	2	US-10-104-047-3209	Sequence 3209, Ap
37	46	38.0	700	2	US-08-931-952-2	Sequence 2, Appli
38	46	38.0	700	2	US-08-272-247-2	Sequence 2, Appli
39	46	38.0	700	4	PCT-US95-08560-2	Sequence 2, Appli
40	45.5	37.6	151	2	US-09-198-452A-815	Sequence 815, App
41	45.5	37.6	205	2	US-09-438-185A-767	Sequence 767, App
42	45	37.2	199	2	US-09-605-703B-2226	Sequence 2226, Ap
43	45	37.2	445	4	PCT-US94-05387-7	Sequence 7, Appli
44	45	37.2	895	1	US-08-123-161A-8	Sequence 8, Appli
45	45	37.2	895	1	US-08-483-278-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1

US-09-290-049A-12
; Sequence 12, Application US/092900049A
; Patent No. 6827936
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; TITLE OF INVENTION: CARRIES
; FILE REFERENCE: 1564.1008-002
; CURRENT APPLICATION NUMBER: US/09/290,049A
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 21
; TYPE: PRT
; ORGANISM: S. downei
US-09-290-049A-12

Query Match 95.0%; Score 115; DB 2; Length 21;

Best Local Similarity 100.0%; Pred. No. 2.2e-11;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNHVSIVEAWSNDTPYLHD 20
DB 2 NNHVSIVEAWSNDTPYLHD 21

RESULT 2

US-09-290-049A-14
; Sequence 14, Application US/092900049A
; Patent No. 6827936
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; TITLE OF INVENTION: CARRIES
; FILE REFERENCE: 1564.1008-002
; CURRENT APPLICATION NUMBER: US/09/290,049A
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19

```
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-290-049A-14

Query Match          95.0%; Score 115; DB 2; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.2e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNHVSIVEAWSNDTPYLHDD 20
   |||||
Db 2 NNHVSIVEAWSNDTPYLHDD 21
   |||||

RESULT 3
US-09-007-999-2
; Sequence 2, Application US/09007999
; Patent No. 6087559
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE REFERENCE: Q356D
; CURRENT APPLICATION NUMBER: US/09/007,999
; CURRENT FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-007-999-2

Query Match          90.9%; Score 110; DB 2; Length 1475;
Best Local Similarity 85.7%; Pred. No. 1.7e-08;
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNHVSIVEAWSNDTPYLHDD 21
   |||||
Db 482 NDHLSILEAWSNDTPYLHDD 502
   |||||

RESULT 4
US-09-210-361-2
; Sequence 2, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE REFERENCE: Q357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-210-361-2

Query Match          90.9%; Score 110; DB 2; Length 1475;
Best Local Similarity 85.7%; Pred. No. 1.7e-08;
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNHVSIVEAWSNDTPYLHDD 21
   |||||
Db 482 NDHLSILEAWSNDTPYLHDD 502
   |||||

RESULT 5
US-09-740-274-2
; Sequence 2, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: Q357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-740-274-2

Query Match          90.9%; Score 110; DB 2; Length 1475;
Best Local Similarity 85.7%; Pred. No. 1.7e-08;
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNHVSIVEAWSNDTPYLHDD 21
   |||||
Db 482 NDHLSILEAWSNDTPYLHDD 502
   |||||

RESULT 6
US-09-290-049A-1
; Sequence 1, Application US/09290049A
; Patent No. 6827936
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; FILE REFERENCE: 1564.1008-002
; CURRENT APPLICATION NUMBER: US/09/290,049A
; CURRENT FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1475
```



```

; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 21
; TYPE: PRT
; ORGANISM: S. mutans
US-09-290-049A-11

```

Query Match 60.3%; Score 73; DB 2; Length 21;
Best Local Similarity 68.4%; Pred. No. 7.9e-05;
Matches 13; Conservative 2; Mismatches 4; Indels

Qy	2	NHVSIVEAWSNDTPYLHD	20
		:	
Db	3	NHLSILEAWSNDPQYNKD	21

```

RESULT 11
US-09-604-957-4
; Sequence 4, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCES: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-604-957-4

```

Query Match 60.3%; Score 73; DB 2; Length 545;
Best Local Similarity 68.4%; Pred. No. 0.0033;
Matches 13; Conservative 2; Mismatches 4; Indels

Qy 2 NHVSIVEAWSNDTPYLHD 20
||:||:||:||:||
Db 77 NHLSILEAWSNDPPOYNKD 95

```

RESULT 12
US-09-995-749A-10
; Sequence 10, Application US/09995749A
; Patent No. 6867026
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LOUBERT
; APPLICANT: RAHAOUTI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: BO43388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-995-749A-10

```

```

Query Match      60.3%; Score 73; DB 2; Length 545;
Best Local Similarity 68.4%; Pred. NO. 0.0033;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy      2 NHVSIVEAWSNDNPTYLHD 20
      ||:||:|||||
Db      77 NHLISLEAWSNDNPQYNKD 95

```

```

RESULT 13
US-09-008-172-2
? Sequence 2, Application US/09008172
? Patent No. 6127602
? GENERAL INFORMATION:
? APPLICANT: Nichols, Scott E.
? TITLE OF INVENTION: Substitutes for Modified Starch and
? TITLE OF INVENTION: Latexes in Paper Manufacture
? FILE REFERENCE: 0358D
? CURRENT APPLICATION NUMBER: US/09/008,172
? CURRENT FILING DATE: 1998-01-16
? EARLIER APPLICATION NUMBER: 08/482,711
? EARLIER FILING DATE: 1995-06-07
? NUMBER OF SEQ ID NOS: 2
? SOFTWARE: FastSEQ for Windows Version 3.0
? SEQ ID NO 2
? LENGTH: 1430
? TYPE: PRT
? ORGANISM: Streptococcus mutans
US-09-008-172-2

```

```

Query Match          60.3%; Score 73; DB 2; Length 1430;
Best Local Similarity 68.4%; Pred. No. 0.0099;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy      2 NHVSIVEAWSNDTPTPYLHD 20
      ||:|||||
Db      497 NHLISLEAWSNDPQYNKD 515

```

```

RESULT 14
US-09-210-361-6
; Sequence 6, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; ; Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210.361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-210-361-6

```

Query Match 60.3%; Score 73; DB 2; Length 1430;
Best Local Similarity 68.4%; Pred. No. 0.0099;

Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 NHVSIVEAWSNDNTPYLHD 20
||:|||||
Db 497 NHLSILEAWSNDNPQYNKD 515

RESULT 15

US-09-740-274-6
; Sequence 6, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-6

Query Match 60.3%; Score 73; DB 2; Length 1430;
Best Local Similarity 68.4%; Pred. No. 0.0099;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 NHVSIVEAWSNDNTPYLHD 20
||:|||||
Db 497 NHLSILEAWSNDNPQYNKD 515

Search completed: February 11, 2006, 19:46:19
Job time : 0.45978 secs

THIS PAGE BLANK (USPTO)


```
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 37
; LENGTH: 1590
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-383-930-37

Query Match          100.0%; Score 121; DB 4; Length 1590;
Best Local Similarity 100.0%; Pred. No. 5.9e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNHVSIVEAWSNDTPLYHDD 21
Db 478 NNHVSIVEAWSNDTPLYHDD 498

RESULT 3
US-10-797-821-37
; Sequence 37, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 37
; LENGTH: 1590
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-797-821-37

Query Match          100.0%; Score 121; DB 5; Length 1590;
Best Local Similarity 100.0%; Pred. No. 5.9e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNHVSIVEAWSNDTPLYHDD 21
Db 478 NNHVSIVEAWSNDTPLYHDD 498

RESULT 4
US-10-797-302-12
; Sequence 12, Application US/10797302
; Publication No. US20050026271A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; TITLE OF INVENTION: Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-019
; CURRENT APPLICATION NUMBER: US/10/797,302
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 37
; LENGTH: 1590
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-797-821-37

Query Match          100.0%; Score 121; DB 5; Length 1590;
Best Local Similarity 100.0%; Pred. No. 5.9e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNHVSIVEAWSNDTPLYHDD 21
Db 478 NNHVSIVEAWSNDTPLYHDD 498

RESULT 5
US-10-797-302-14
; Sequence 14, Application US/10797302
; Publication No. US20050026271A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-019
; CURRENT APPLICATION NUMBER: US/10/797,302
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-02
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 14
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-797-302-14

Query Match          95.0%; Score 115; DB 5; Length 21;
Best Local Similarity 100.0%; Pred. No. 3.8e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNHVSIVEAWSNDTPLYLHD 20
Db 2 NNHVSIVEAWSNDTPLYLHD 21

RESULT 6
US-09-740-274-2
; Sequence 2, Application US/09740274
; Patent No. US20020031826A1
```

```

: GENERAL INFORMATION:
: APPLICANT: Smith, Daniel J.
: APPLICANT: Taubman, Martin A.
: TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
: FILE REFERENCE: 25669-020
: CURRENT APPLICATION NUMBER: US/10/797,821
: CURRENT FILING DATE: 2004-03-09
: PRIOR APPLICATION NUMBER: 10/383,930
: PRIOR FILING DATE: 2003-03-07
: PRIOR APPLICATION NUMBER: 60/363,209
: PRIOR FILING DATE: 2002-03-07
: PRIOR APPLICATION NUMBER: 60/402,483
: PRIOR FILING DATE: 2002-08-08
: PRIOR APPLICATION NUMBER: 09/290,049
: PRIOR FILING DATE: 1999-04-12
: PRIOR APPLICATION NUMBER: 60/081,550
: PRIOR FILING DATE: 1998-04-13
: PRIOR APPLICATION NUMBER: 60/115,142
: PRIOR FILING DATE: 1999-01-08
: NUMBER OF SEQ ID NOS: 45
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 34
: LENGTH: 1475
: TYPE: PRT
: ORGANISM: Streptococcus mutans
US-10-797-821-34

```

Query Match 90.9%; Score 110; DB 5; Length 1475;
Best Local Similarity 85.7%; Pred. No. 2.4e-07;
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY	I NNRVSIWEAWSNDNDFITRDD ZI
	: : : : : : :
D6	482 NDHLSILEAWSNDNTPYLHDD 502

```

RESULT 9
US-10-797-302-1
; Sequence 1, Application US/10797302
; Publication No. US20050026271A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-019
; CURRENT APPLICATION NUMBER: US/10/797,302
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-02
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: EAW peptide
US-10-797-302-1

```

NUMBER OF SEQ ID NOS: 21
SOFTWARE: PatentIn version 3.2

```

; LIFE: PRI
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: EAW peptide
US-10-797-302-1

```

Query Match 86.0%; Score 104; DB 5; Length 21;
Best Local Similarity 85.0%; Pred. No. 1.7e-08;
Matches 17; Conservative 3; Mismatches 0; Indels

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - proteih search, using sw model

Run on: February 11, 2006, 20:42:47 ; Search time 0.119502 Seconds
(without alignments)
2306.008 Million cell updates/sec

Title: US-10-797-821-42

Perfect score: 121

Sequence: 1 NHVSVIVEAWSNDTPYLHDD 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 97014 seqs, 13122538 residues

Total number of hits satisfying chosen parameters: 97014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : . Pubmed Applications AA_New.*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB_PEP.*
2: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB_PEP.*
3: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB_PEP.*
4: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB_PEP.*
5: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB_PEP.*
6: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB_PEP.*
7: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB_PEP.*
8: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB_PEP.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	49.5	40.9	1604	7 US-11-037-243-73	Sequence 73, Appl
2	47	38.8	1267	7 US-11-109-156-35	Sequence 35, Appl
3	45	37.2	895	7 US-11-150-406-2	Sequence 2, Appli
4	45	37.2	920	6 US-10-821-234-1129	Sequence 1129, Ap
5	45	37.2	1366	7 US-11-169-041-174	Sequence 174, App
6	44	36.4	435	7 US-11-010-239-113	Sequence 113, App
7	43.5	36.0	443	7 US-11-112-882-29	Sequence 113, App
8	43	35.5	318	6 US-10-873-528-192	Sequence 192, App
9	43	35.5	318	7 US-11-052-554A-233	Sequence 233, App
10	43	35.5	617	6 US-10-995-561-890	Sequence 890, App
11	42.5	35.1	654	6 US-10-528-031-1	Sequence 1, Appli
12	42	34.7	502	7 US-11-098-686-10593	Sequence 10593, A
13	42	34.7	1627	7 US-11-052-554A-124	Sequence 124, App
14	41	33.9	746	7 US-11-052-554A-57	Sequence 57, Appl
15	41	33.9	1432	6 US-10-510-386-218	Sequence 218, App
16	41	33.9	1607	7 US-11-098-686-10178	Sequence 10178, A
17	40.5	33.5	155	6 US-10-793-626-308	Sequence 308, App
18	40.5	33.5	335	7 US-11-183-664-2	Sequence 2, Appli
19	40.5	33.5	353	7 US-11-024-959-473	Sequence 473, App
20	40.5	33.5	413	7 US-11-052-554A-308	Sequence 308, App
21	40.5	33.5	534	7 US-11-167-856-24	Sequence 24, Appl
22	40.5	33.5	1238	7 US-11-078-735-21	Sequence 21, Appl
23	40.5	33.5	1238	7 US-11-050-346-66	Sequence 66, Appl
24	40.5	33.5	1238	7 US-11-103-077-21	Sequence 21, Appl
25	40	33.1	839	7 US-11-076-431-2	Sequence 2, Appli

Sequence 4, Appli
Sequence 6, Appli
Sequence 8, Appli
Sequence 2, Appli
Sequence 3490, Ap
Sequence 10464, A
Sequence 1416, Ap
Sequence 288, App
Sequence 42, Appl
Sequence 10530, A
Sequence 18, Appl
Sequence 33, Appl
Sequence 4834, Ap
Sequence 4, Appli
Sequence 1138, Ap
Sequence 10214, A
Sequence 640, App
Sequence 7694, Ap
Sequence 138, App

26 40 33.1 839 7 US-11-076-431-4
27 40 33.1 839 7 US-11-076-431-6
28 40 33.1 839 7 US-11-076-431-8
29 40 33.1 839 7 US-11-005-216-2
30 39.5 32.6 98 6 US-10-467-657-3490
31 39.5 32.6 655 7 US-11-098-686-10464
32 39 32.2 220 6 US-10-793-626-1416
33 39 32.2 299 7 US-11-156-084-288
34 39 32.2 361 7 US-11-012-762-42
35 39 32.2 455 7 US-11-098-686-10530
36 39 32.2 546 6 US-10-661-966-18
37 38.5 31.8 236 6 US-10-878-556A-33
38 38.5 31.8 293 6 US-10-467-657-4834
39 38.5 31.8 432 7 US-11-140-417-4
40 38 31.4 186 7 US-11-182-016-40
41 38 31.4 266 7 US-11-055-822-1138
42 38 31.4 289 7 US-11-098-686-10214
43 38 31.4 324 6 US-10-793-626-640
44 38 31.4 436 6 US-10-467-657-7694
45 38 31.4 445 6 US-10-878-556A-138

ALIGNMENTS

RESULT 1

US-11-037-243-73

; Sequence 73, Application US/11037243

; Publication No. US20050287546A1

; GENERAL INFORMATION:

; APPLICANT: PLOWMAN, GREGORY

; APPLICANT: WHYTE, DAVID

; APPLICANT: CAENEPEEL, SEAN

; APPLICANT: CHARYDCZAK, GLEN

; APPLICANT: MANNING, GERARD

; APPLICANT: SUDARSANAN, SUCHA

; TITLE OF INVENTION: NOVEL PROTEASES

; FILE REFERENCE: 038602/1214

; CURRENT APPLICATION NUMBER: US/11/037,243

; PRIOR FILING DATE: 2005-05-26

; PRIOR APPLICATION NUMBER: US/09/888,615

; PRIOR FILING DATE: 2001-06-26

; PRIOR APPLICATION NUMBER: 60/214,047

; NUMBER OF SEQ ID NOS: 150

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 73

; LENGTH: 1604

; TYPE: PRT

; ORGANISM: Homo sapiens

US-11-037-243-73

Query Match 40.9%; Score 49.5; DB 7; Length 1604;
Best Local Similarity 47.6%; Pred. No. 11;
Matches 10; Conservative 3; Mismatches 5; Indels 3; Gaps 1;

QY 4 VSVIVEAWSN---DTPYLHDD 21

Db 294 VALLEVKDNRDTDIPELHMD 314

RESULT 2

US-11-109-156-35

; Sequence 35, Application US/1109156

; Publication No. US20050250144A1

; GENERAL INFORMATION:

; APPLICANT: Toshio Ota

; APPLICANT: Takao Isogai

; APPLICANT: Tetsumi Nishikawa

; APPLICANT: Koji Hayashi

; APPLICANT: Kaoru Otsuka

; APPLICANT: Jun-ichi Yamamoto

; APPLICANT: Shizuko Ishii


```
; APPLICANT: Tomoyasu Sugiyama
; APPLICANT: Ai Wakamatsu
; APPLICANT: Keiichi Nagai
; APPLICANT: Tetsuji Otsuki
; APPLICANT: Shin-Ichi Funahashi
; APPLICANT: Chiaki Senoo
; APPLICANT: Jun-Ichi Nezu
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEIN KINASE/PROTEIN
; FILE REFERENCE: 06501-099002
; CURRENT APPLICATION NUMBER: US/11/109,156
; CURRENT FILING DATE: 2005-04-19
; PRIOR APPLICATION NUMBER: US/10/060,065
; PRIOR FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: PCT/JP00/05061
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: US 60/159,590
; PRIOR FILING DATE: 1999-10-18
; PRIOR APPLICATION NUMBER: US 60/183,322
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: JP 11-248036
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: JP 2000-118776
; PRIOR FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: JP 2000-183767
; PRIOR FILING DATE: 2000-05-02
; PRIOR APPLICATION NUMBER: JP 2000-241899
; PRIOR FILING DATE: 2000-06-09
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35
; LENGTH: 1267
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-109-156-35

Query Match 38.8%; Score 47; DB 7; Length 1267;
Best Local Similarity 47.6%; Pred. No. 20;
Matches 10; Conservative 3; Mismatches 6; Indels 2; Gaps 1;

QY 3 HVSIVE--AWSNDTPYLHDD 21
Db 1149 HSHLNFTAWPDHDTSPQDD 1169

RESULT 3
US-11-150-406-2
; Sequence 2, Application US/11/150406
; Publication No. US20050250164A1
; GENERAL INFORMATION:
; APPLICANT: Muschler, John L
; APPLICANT: Bissell, Mina J
; TITLE OF INVENTION: Design of Novel Assays Based on the Newly Found Role of
; TITLE OF INVENTION: Dystroglycan and Alpha-Dystroglycan Proteolysis in Tumor Cell
; TITLE OF INVENTION: Growth
; FILE REFERENCE: IB-1398A
; CURRENT APPLICATION NUMBER: US/11/150,406
; CURRENT FILING DATE: 2005-06-09
; PRIOR APPLICATION NUMBER: 60/151,766
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: 09/652,493
; PRIOR FILING DATE: 2000-08-31
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2
; LENGTH: 895
; TYPE: PRT
; ORGANISM: homo sapiens
US-11-150-406-2

Query Match 37.2%; Score 45; DB 7; Length 895;
Best Local Similarity 35.7%; Pred. No. 29;
Matches 10; Conservative 4; Mismatches 6; Indels 6; Gaps 1;
```

```
QY 2 NHVSIVEAW-----SDNDTPYLHDD 21
Db 501 NHIDRVDAWVGTYFEVKIPSDTFYDHD 528

RESULT 4
US-10-821-234-1129
; Sequence 1129, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt SEQ_genes Version 1.0
; SEQ ID NO 1129
; LENGTH: 920
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1129

Query Match 37.2%; Score 45; DB 6; Length 920;
Best Local Similarity 35.7%; Pred. No. 29;
Matches 10; Conservative 4; Mismatches 6; Indels 8; Gaps 1;

QY 2 NHVSIVEAW-----SDNDTPYLHDD 21
Db 526 NHIDRVDAWVGTYFEVKIPSDTFYDHD 553

RESULT 5
US-11-169-041-174
; Sequence 174, Application US/11169041
; Publication No. US20060019284A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES FOR PREDICTING ACTIVITY OF
; TITLE OF INVENTION: COMPOUNDS THAT INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE
; TITLE OF INVENTION: KINASES AND/OR PROTEIN TYROSINE KINASE PATHWAYS IN LUNG CANCER
; TITLE OF INVENTION: CELLS
; FILE REFERENCE: 10001 NP
; CURRENT APPLICATION NUMBER: US/11/169,041
; CURRENT FILING DATE: 2005-06-28
; PRIOR APPLICATION NUMBER: 60/584,405
; PRIOR FILING DATE: 2004-06-30
; NUMBER OF SEQ ID NOS: 527
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 174
; LENGTH: 1366
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-169-041-174

Query Match 37.2%; Score 45; DB 7; Length 1366;
Best Local Similarity 47.4%; Pred. No. 44;
Matches 9; Conservative 1; Mismatches 7; Indels 2; Gaps 1;

QY 2 NHVSIVEAWSNDTPYLHD 20
Db 130 NOVSTI--WDDNPAFSTHD 146

RESULT 6
US-11-010-239-113
; Sequence 113, Application US/11010239
```

```
; Publication No. US20060015970A1
; GENERAL INFORMATION:
; APPLICANT: ROGER PENNELL
; APPLICANT: JACK OKAMURO
; APPLICANT: RICHARD SCHNEEBERGER
; APPLICANT: YI-WEN FANG
; APPLICANT: SHING KWOK
; APPLICANT: DIANE JOFUKU
; APPLICANT: EDWARD A. KIEGLE
; APPLICANT: JONATHAN DONSON
; APPLICANT: NESTOR APUYA
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES AND POLYPEPTIDES ENCODED THEREBY USEFUL FOR
; TITLE OF INVENTION: MODIFYING PLANT CHARACTERISTICS
; FILE REFERENCE: 2750-1585PUS2
; CURRENT APPLICATION NUMBER: US/11/010,239
; PRIOR FILING DATE: 2004-12-09
; PRIOR APPLICATION NUMBER: US 60/529,352
; PRIOR FILING DATE: 2003-12-12
; NUMBER OF SEQ ID NOS: 133
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 113
; LENGTH: 435
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(435)
; OTHER INFORMATION: 12600234_protein_ID_12600235
US-11-010-239-113

Query Match 36.4%; Score 44; DB 7; Length 435;
Best Local Similarity 50.0%; Pred. No. 19;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 6 IVEAWSNDTPY 17
Db 344 VLEAWSDKTSFP 355

RESULT 7
US-11-112-882-29
; Sequence 29, Application US/11112882
; Publication No. US20050273885A1
; GENERAL INFORMATION:
; APPLICANT: Commonwealth Scientific and Industrial Research Organisation
; TITLE OF INVENTION: Synthesis of Long-Chain Polyunsaturated Fatty Acids in Recombinant
; FILE REFERENCE: 503244
; CURRENT APPLICATION NUMBER: US/11/112,882
; CURRENT FILING DATE: 2005-04-22
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 29
; LENGTH: 443
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-11-112-882-29

Query Match 36.0%; Score 43.5; DB 7; Length 443;
Best Local Similarity 45.5%; Pred. No. 23;
Matches 10; Conservative 3; Mismatches 8; Indels 1; Gaps 1;

QY 1 NNHVSIVEAW-SDNDTPYLYHDD 21
Db 398 NACVKIVKWKCKENNLPLYLVD 419

RESULT 8
US-10-873-528-192
; Sequence 192, Application US/10873528
; Publication No. US20050276814A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
```

```
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129WO
; CURRENT APPLICATION NUMBER: US/10/873,528
; CURRENT FILING DATE: 2004-06-23
; PRIOR APPLICATION NUMBER: US/09/769,787
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 192
; LENGTH: 318
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-873-528-192

Query Match 35.5%; Score 43; DB 6; Length 318;
Best Local Similarity 40.9%; Pred. No. 20;
Matches 9; Conservative 3; Mismatches 8; Indels 2; Gaps 1;

QY 2 NHVSIVEAWSNDTP--YLHDD 21
Db 170 NGLTETGQKNDTGYWVHSD 191

RESULT 9
US-11-052-554A-233
; Sequence 233, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 233
; LENGTH: 318
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae R6
US-11-052-554A-233

Query Match 35.5%; Score 43; DB 7; Length 318;
Best Local Similarity 40.9%; Pred. No. 20;
Matches 9; Conservative 3; Mismatches 8; Indels 2; Gaps 1;

QY 2 NHVSIVEAWSNDTP--YLHDD 21
Db 170 NGLTETGQKNDTGYWVHSD 191

RESULT 10
US-10-995-561-890
; Sequence 890, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 890
; LENGTH: 617
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-890

Query Match          35.5%; Score 43; DB 6; Length 617;
Best Local Similarity 38.9%; Pred. No. 39;
Matches 7; Conservative 3; Mismatches 0; Gaps 0; Indels 8;

QY 3 HVSIVEAWSNDNTPYLHD 20
Db 97 NADLVVLWTDGDTATPAD 114

RESULT 11
US-10-528-031-1
; Sequence 1, Application US/10528031
; Publication No. US20050262577A1
; GENERAL INFORMATION:
; APPLICANT: ORIDIS BIOMED Forschungs- und Entwicklungs GmbH
; APPLICANT: Guelly, Christian
; APPLICANT: Buck, Charles R.
; APPLICANT: Zatloukal, Kurt
; TITLE OF INVENTION: Polypeptides and nucleic acids encoding these and their use for
; TITLE OF INVENTION: prevention, diagnosis or treatment of liver disorders and epithe
; FILE REFERENCE: Oridis Biomed
; CURRENT APPLICATION NUMBER: US/10/528,031
; CURRENT FILING DATE: 2005-03-16
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 654
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-528-031-1

Query Match          35.1%; Score 42.5; DB 6; Length 654;
Best Local Similarity 64.3%; Pred. No. 49;
Matches 9; Conservative 2; Mismatches 1; Gaps 1; Indels 2;

QY 5 SIV-EAWSNDNTPY 17
Db 251 SIVPOSWSNPPTY 264

RESULT 12
US-11-098-686-10593
; Sequence 10593, Application US/11098686
; Publication No. US20060024696A1
; GENERAL INFORMATION:
; APPLICANT: Kapur, Vivek and Gebhart, Connie J.
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
; TITLE OF INVENTION: FROM LAWSONIA INTRACELLULARIS AND METHODS OF USING
; FILE REFERENCE: 09531-128001
; CURRENT APPLICATION NUMBER: US/11/098,686
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: PCT/US03/31318
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: US 60/416,395
; PRIOR FILING DATE: 2002-10-04
; NUMBER OF SEQ ID NOS: 11433
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10593
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Lawsonia intracellularis
US-11-098-686-10593

Query Match          34.7%; Score 42; DB 7; Length 502;
Best Local Similarity 37.5%; Pred. No. 44;
Matches 6; Conservative 5; Mismatches 0; Gaps 0; Indels 5;
```

```
QY 1 NNHVSIVEAWSNDNTP 16
Db 186 NPYFALLESWKNGTTP 201

RESULT 13
US-11-052-554A-124
; Sequence 124, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 124
; LENGTH: 1627
; TYPE: PRT
; ORGANISM: Mycoplasma pneumoniae
US-11-052-554A-124

Query Match          34.7%; Score 42; DB 7; Length 1627;
Best Local Similarity 39.3%; Pred. No. 1.5e+02;
Matches 11; Conservative 2; Mismatches 3; Indels 12; Gaps 1;

QY 4 VSIVEA-----WSNDNTPYLH 19
Db 1214 VSVVEASAYKNTSSGQTQSTNSPYLH 1241

RESULT 14
US-11-052-554A-57
; Sequence 57, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 57
; LENGTH: 746
; TYPE: PRT
; ORGANISM: Escherichia coli 0157:H7
US-11-052-554A-57

Query Match          33.9%; Score 41; DB 7; Length 746;
Best Local Similarity 43.8%; Pred. No. 94;
Matches 7; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 6 IVEAWSNDNTPYLHDD 21
Db 19 VAQAQEPDTTPVSHDD 34

RESULT 15
US-10-510-386-218
; Sequence 218, Application US/10510386
```

; Publication No. US20050244922A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Jens Tonne
; APPLICANT: Clausen, Ib Groth
; APPLICANT: Jorgensen, Steen Troels
; APPLICANT: Olsen, Peter Bjarke
; APPLICANT: Rasmussen, Michael Dolberg
; TITLE OF INVENTION: Improved Bacillus Host Cell
; FILE REFERENCE: 10294.204-US
; CURRENT APPLICATION NUMBER: US/10/510,386
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 248
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 218
; LENGTH: 1432
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-10-510-386-218

Query Match 33.9%; Score 41; DB 6; Length 1432;
Best Local Similarity 42.1%; Pred. No. 1.8e+02;
Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

Qy 2 NHVSIVEANSNDTPYLHD 20
Db 533 NHQETIYVSGSETPLKAD 551

Search completed: February 11, 2006, 20:59:02
Job time : 0.119502 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 19:15:50 ; Search time 0.433449 Seconds
(without alignments)
4661.567 Million cell updates/sec

Title: US-10-797-821-42
Perfect score: 121
Sequence: 1 NNHVSIVEAWSNDTPYLHDD 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	121	100.0	1592	2 A38175	glucosyltransferas
2	110	90.9	1475	2 B33135	gtfB protein precu
3	101	83.5	1375	2 JT0345	dextranucrase (EC
4	73	60.3	1431	2 A45866	dextranucrase (EC
5	66	54.5	1365	2 A41483	glucosyltransferas
6	64	52.9	1508	2 T31098	probable dextranu
7	60	49.6	1449	2 T30857	glucosyltransferas
8	60	49.6	1449	2 T30552	glucosyltransferas
9	57	47.1	1577	2 T30858	glucosyltransferas
10	56	46.3	1518	2 A44811	glucosyltransferas
11	55	45.5	1599	2 S22737	glucosyltransferas
12	53	43.8	175	2 C86205	glucosyltransferas
13	49	40.5	396	2 T35024	probable glutathio
14	48	39.7	331	2 B48445	glyceraldhyde-3-p
15	47	38.8	107	2 S40290	protein-tyrosine-p
16	47	38.8	336	1 Q3B40	BglF2 protein - nu
17	47	38.8	1090	2 AG1749	glucosidase homolo
18	47	38.8	2294	2 I67630	protein-tyrosine-p
19	47	38.8	2450	2 S71625	protein-tyrosine-p
20	47	38.8	2466	2 I67629	protein-tyrosine-p
21	47	38.8	2490	1 A54971	protein-tyrosine-p
22	46	38.0	313	2 S59448	hypothetical prote
23	46	38.0	490	2 H70538	probable ppdK prot
24	46	38.0	601	2 E87028	pyruvate, phosphat
25	46	38.0	947	2 E86362	hypothetical prote
26	46	38.0	1100	2 AF1460	alpha-xylosidase a
27	46	38.0	1100	2 AG1097	alpha-xylosidase a
28	45.5	37.6	194	2 H72037	conserved hypothet
29	45.5	37.6	194	2 C86586	CT647 hypothetical

30	45.5	37.6	418	2 D90506	4-aminobutyrate am
31	45	37.2	337	2 C64233	glyceraldhyde-3-p
32	45	37.2	347	2 T48610	hypothetical prote
33	45	37.2	524	2 D82220	conserved hypotet
34	45	37.2	525	2 T40088	RhoGEF domain cont
35	45	37.2	740	2 T15197	hypothetical prote
36	45	37.2	895	2 I54343	dystroglycan - hum
37	45	37.2	1131	2 T16217	hypothetical prote
38	45	37.2	1384	2 T02748	hypothetical prote
39	44.5	36.8	835	2 B64689	site-specific DNA-
40	44	36.4	316	1 A42935	N-acetylmutamoyl-L
41	44	36.4	735	2 T19065	probable protein k
42	44	36.4	1313	2 T29027	hypothetical prote
43	43.5	36.0	279	1 SNPF5K	proteasome endopep
44	43.5	36.0	491	1 D64947	glucose-6-phosphat
45	43.5	36.0	491	2 AB0742	glucose-6-phosphat

ALIGNMENTS

RESULT 1

A38175
glucosyltransferase precursor - Streptococcus sobrinus
C;Species: Streptococcus sobrinus
C;Date: 28-Aug-1992 #sequence_revision 28-Aug-1992 #text_change 15-Mar-2004
C;Accession: A38175
R;Abo, H.; Mtsamura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.
J. Bacteriol. 173, 989-996, 1991
A;Title: Peptide sequences for sucrose splitting and glucan binding within Streptococcus
A;Reference number: A38175; MUID:91123227; PMID:1704006
A;Accession: A38175
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1592 <ABO>
A;Cross-references: UNIPARC:UPI000012BCB2; GB:D90213; NID:G217032; PIDN:BAA14241.1; PID
F;1093-1112/Domain: cpl repeat homology <CP1>
F;1222-1241/Domain: cpl repeat homology <CP2>
F;1287-1306/Domain: cpl repeat homology <CP3>
F;1330-1351/Domain: cpl repeat homology <CP4>
F;1352-1371/Domain: cpl repeat homology <CP5>
F;1402-1420/Domain: cpl repeat homology <CP6>
F;1465-1484/Domain: cpl repeat homology <CP7>
F;1513-1532/Domain: cpl repeat homology <CP8>

Query Match 100.0%; Score 121; DB 2; Length 1592;
Best Local Similarity 100.0%; Pred.No. 4.3e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNHVSIVEAWSNDTPYLHDD 21
|||||
Db 478 NNHVSIVEAWSNDTPYLHDD 498

RESULT 2

B33135
gtfB protein precursor - Streptococcus mutans
C;Species: Streptococcus mutans
C;Date: 23-Oct-1990 #sequence_revision 23-Oct-1990 #text_change 09-Jul-2004
C;Accession: B33135; A33128
R;Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
J. Bacteriol. 169, 4263-4270, 1987
A;Title: Sequence analysis of the gtfB gene from Streptococcus mutans.
A;Reference number: A33135; MUID:87308013; PMID:3040685
A;Accession: B33135
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1475 <SHI>
A;Cross-references: UNIPROT:P08987; UNIPARC:UPI000014D972; GB:M17361; NID:G153639; PIDN
R;Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
submitted to the Protein Sequence Database, September 1990
A;Reference number: A33128
A;Accession: A33128

RESULT 9
T30858
glucosyltransferase - Streptococcus salivarius
C:Species: Streptococcus salivarius
C/Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C/Accession: T30858
R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.
Infect. Immun. 63. 609-621. 1995

A;Accession: S22737
A;Molecule type: DNA
A;Residues: 1-1599 <JAC>
A;Cross-references: UNIPROT:Q00599; UNIPARC:UPI00000BEF34; EMBL:Z11872; NID:G47530; PID:
R;Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.
J. Gen. Microbiol. 137, 2577-2593, 1991
A;Title: Molecular characterization of a cluster of at least two glucosyltransferase ge
A;Reference number: A44811, MUID:92148377; PMID:1638391
A;Accession: S28810
A;Molecule type: DNA
A;Residues: 1-51 <GIF>
A;Cross-references: UNIPARC:UPI000017027C; EMBL:Z11873
C;Genetics:
A;Gene: gtfK
C;Keywords: glucosyltransferase; hexosyltransferase

F:1456-1475/Domain: cpl repeat homology <CPR>

```

Query Match          45.58;   Score 55;   DB 2;   Length 1599;
Best Local Similarity 50.04;   Pred. No. 6.3;
Matches 9; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Qy      3 HVSGIVEAWSNDPTPLHLD 20
        :|||:||| || |:::
Db      494 NISILEAWSHNDPPVYNE 511

RESULT 12
C86205
hypothetical protein [imported] - Arabidopsis thaliana
C;Species:Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C;Accession: C86205
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.B.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: AB6141; MUID: 21016719; PMID:11130712
A;Accession: C86205
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-175 <STO>
A;Cross-references: UNIPROT:Q9LMJ8; UNIPARC:UPI000000AJ10C0; GB:ABE005172; NID:GB954041; PT
C;Genetics:
A;Map position: 1

```

```

RESULT 14
B48445
glyceraldehyde-3-phosphate dehydrogenase (phosphorylating) (EC 1.2.1.12) - Leishmania m
C/Species: Leishmania mexicana
C/Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C/Accession: B48445; S25142
R:Hannaert, V.; Blaauw, M.; Kohl, L.; Allert, S.; Oppendoes, F.R.; Michels, P.A.M.
Mol. Biochem. Parasitol. 55, 115-126, 1992
A/Title: Molecular analysis of the cytosolic and glycosomal glyceraldehyde-3-phosphate (
A/Reference number: A48445; MUID:93063042; PMID:1435864
A/Accession: B48445
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-331 <HAN>
A/Cross-references: UNIPROT:Q01558; UNIPARC:UPI0000016BF19; EMBL:X65220; NID:g9552; PIDN
C/Superfamily: glyceraldehyde-3-phosphate dehydrogenase
C/Keywords: oxidoreductase

Query Match      39.7%; Score 48; DB 2; Length 331;
Best Local Similarity 42.1%; Pred. No. 12;
Matches      8; Conservative      4; Mismatches      7; Indels      0; Gaps      0;

Qy      1 NNHVSIVEAWSNDTPYLH 19
      ||| : ||| : ||| : |||
Db      301 NDHFVXLVSYDNETGYSH 319

RESULT 15
S40290
protein-tyrosine-phosphatase (EC 3.1.3.48), nonreceptor type 13 - mouse (fragment)
C/Species: Mus musculus (house mouse)
C/Date: 06-Mar-1994 #sequence_revision 15-Mar-1996 #text_change 05-Oct-2004
C/Accession: S40290
R:Hendriks, W.; Brugman, C.; Zeeuwen, P.; Schepens, J.; Wieringa, B.
submitted to the EMBL Data Library, June 1993
A/Description: Assessment of the expression levels of murine protein-tyrosine phosphatase
A/Reference number: S40280
A/Accession: S40290
A/Molecule type: mRNA
A/Residues: 1-107 <HEN>
A/Cross-references: UNIPROT:Q64512; UNIPARC:UPI0000016CFDA; EMBL:Z23059; NID:g438155; PFI
C/Superfamily: protein-tyrosine phosphatase, non-receptor type 13; GLGF domain homology
C/Keywords: phosphoric monoester hydrolase; tyrosine-specific phosphatase
F:1-107/Domain: protein-tyrosine-phosphatase homology (fragment) <PTP>

Query Match      38.8%; Score 47; DB 2; Length 107;
Best Local Similarity 47.6%; Pred. No. 4.3;
Matches      10; Conservative      3; Mismatches      6; Indels      2; Gaps      1;

Qy      3 HVSIVE--AWSNDTPYLHDD 21
      ||| : ||| : ||| : |||
Db      68 HISHLNFTAWPDHDTSPQDD 88

Search completed: February 11, 2006, 19:42:59
Job time : 1.43345 secs

```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 18:57:35 ; Search time 2.83362 Seconds
(without alignments)
5228.676 Million cell updates/sec

Title: US-10-797-821-42
Perfect score: 121
Sequence: 1 NNHVSIVEAWSNDTPYLHDD 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot_05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	121	100.0	1590	2	Q55263_9STRE
2	121	100.0	1590	2	Q59983_9STRE
3	121	100.0	1592	1	GTF2_STRDO
4	121	100.0	1597	1	GTF1_STRDO
5	110	90.9	1476	1	GTF2_STRMU
6	101	83.5	1455	1	GTF2_STRMU
7	76	62.8	1016	2	Q9LCJ7_LEUME
8	76	62.8	1454	2	Q69A94_LEUME
9	73	60.3	1462	1	GTFD_STRMU
10	68	56.2	2835	2	Q8G9Q2_LEUME
11	67	55.4	1575	2	Q9LCH3_STROR
12	67	55.4	1577	2	Q54178_STRGN
13	66	54.5	1338	2	Q9WXJ4_9STRE
14	66	54.5	1365	1	GTF5_STRDO
15	64	52.9	1330	2	Q84CN4_LEUME
16	64	52.9	1477	2	Q9L466_LEUME
17	64	52.9	1508	2	Q9EZH5_LEUME
18	64	52.9	1508	2	Q52224_LEUME
19	61	50.4	1506	2	Q56CX8_9STRE
20	61	50.4	1522	2	Q6TXV4_LEUME
21	61	50.4	1522	2	Q8KRE1_LEUME
22	61	50.4	1527	2	Q9ZAR4_LEUME
23	61	50.4	1554	2	Q8KZL5_9STRE
24	60	49.6	374	2	Q6ZXI9_9LACO
25	60	49.6	1449	2	Q68542_STRSL
26	60	49.6	1449	2	Q55264_STRSL
27	60	49.6	1463	2	Q5SBM6_LACFE
28	60	49.6	1512	2	Q9WXJ5_9STRE
29	58	47.9	1561	2	Q5SBM8_9LACO
30	58	47.9	1595	2	Q5SBM3_LACSK
31	57	47.1	1577	2	Q55265_STRSL

32	56	46.3	1518	2	Q00600_STRSL
33	55	45.5	1442	2	Q7R2S1_GIALA
34	55	45.5	1599	2	Q00599_STRSL
35	55	45.5	2493	2	Q6CMV4_KLJLA
36	53	43.8	175	2	Q9LNU8_ARATH
37	53	43.8	195	2	Q8LDM8_ARATH
38	52.5	43.4	509	2	Q4WBE0_ASPFU
39	51.5	42.6	767	2	Q4IEN6_GIBZE
40	51	42.1	523	2	Q6CI50_YARLI
41	51	42.1	892	2	Q55VC2_CRYNE
42	51	42.1	892	2	Q5KKW3_CRYNE
43	51	42.1	2057	2	Q9RE05_LEUME
44	50	41.3	395	2	Q4FUX5_9GAMM
45	50	41.3	560	2	Q73WJ6_MYCPA

ALIGNMENTS

RESULT 1
Q55263_9STRE PRELIMINARY; PRT; 1590 AA.
AC Q55263; 9STRE PRELIMINARY; PRT; 1590 AA.
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE GTF-I.
GN Name=Glucosyltransferase;
OS Streptococcus sobrinus;
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OX Streptococcus.
NCBI_TaxID=1310;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 33478;
RA Sato S.;
RT "DNA and amino-acid sequences of water-insoluble-glucan synthetase produced from Streptococcus sobrinus ATCC 33478.";
RL Kagoshima Daigaku Shigakubu Kiyo 16:23-29 (1996).
RN [2]
RP PROTEIN SEQUENCE.
RX MEDLINE=91224988; PubMed=1827439;
RA Mooser G., Hefta S.A., Faxton R.J., Shively J.E., Lee T.D.;
RT "Isolation and sequence of an active-site peptide containing a catalytic aspartic acid from two Streptococcus sobrinus alpha-glucosyltransferases";
RL J. Biol. Chem. 266:8916-8922 (1991).
DR EMBL; D63570; BAA09792.1; -; Genomic_DNA.
DR PIR; A39841; A39841.
DR HSP; P06653; 1GVN.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 3.
DR Pfam; PF02324; Glyco_hydro_70; 1.
SQ SEQUENCE 1590 AA; 176058 MW; 9DF7A3F2C6E4FD43 CRC64;

Query Match 100.0%; Score 121; DB 2; Length 1590;
Best Local Similarity 100.0%; Pred. No. 1.9e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNHVSIVEAWSNDTPYLHDD 21

Db 478 NNHVSIVEAWSNDTPYLHDD 498

RESULT 2

Q59983_9STRE PRELIMINARY; PRT; 1590 AA.
ID Q59983_9STRE PRELIMINARY; PRT; 1590 AA.
AC Q59983;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)

```

DE Glucosyltransferase-I precursor (EC 2.4.1.5).
GN Name=gtf1;
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OC NCBI_TaxID=1310;
RN [1]_TaxID=1310;
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OMZ176;
RX MEDLINE=94146405; PubMed=8312602;
RA Sato S., Inoue M., Handa N., Aizawa Y., Isobe Y., Katayama T.;
RT "DNA sequence of the glucosyltransferase gene of serotype d
RT Streptococcus sobrinus.";
RL DNA Seq. 4:19-27(1993).
RN [2]
RP PROTEIN SEQUENCE.
RX MEDLINE=91224988; PubMed=1827439;
RA Mooser G., Hefta S.A., Paxton R.J., Shively J.E., Lee T.D.;
RT "Isolation and sequence of an active-site peptide containing a
RT catalytic aspartic acid from two Streptococcus sobrinus alpha-
RT glucosyltransferases.";
RL J. Biol. Chem. 266:8916-8922(1991).
DR EMBL; D13858; BAA02976.1; -; Genomic_DNA.
DR PIR; A39841; A39841.
DR HSSP; P06653; 1HCX.
DR GO; GO:0047849; F:dextranase activity; IEA.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_I; 2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
DR Glycosyltransferase; Signal; Transferase.
FT SIGNAL 1 38
FT CHAIN 39 1590 Glucosyltransferase-I.
SQ SEQUENCE 1590 AA; 175956 MW; C3C83A57CF3C2B0E CRC64;

Query Match 100.0%; Score 121; DB 2; Length 1590;
Best Local Similarity 100.0%; Pred. No. 1.9e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNHVSIVEAWSNDTPYLHDD 21
Db 478 NNHVSIVEAWSNDTPYLHDD 498

RESULT 3
GTF2_STRDO
ID_GTF2_STRDO STANDARD; PRT; 1592 AA.
AC P27470;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranase)
DE (Sucrose 6-glucosyltransferase).
OS Streptococcus downei (Streptococcus sobrinus).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OC NCBI_TaxID=1317;
RN [1]_TaxID=1317;
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC STRAIN=6715 / Serotype G;
RX MEDLINE=91123227; PubMed=1704006;
RA Abo H., Matsumura T., Kodama T., Ohta H., Fukui K., Kato K.,
RA Kagawa H.;
RT "Peptide sequences for sucrose splitting and glucan binding within
RT Streptococcus sobrinus glucosyltransferase (water-insoluble glucan
RT synthetase).";
RL J. Bacteriol. 173:989-996(1991).
CC -!- FUNCTION: Production of extracellular glucans, that are thought to
CC play a key role in the development of the dental plaque because of
CC their ability to adhere to smooth surfaces and mediate the
CC aggregation of bacterial cells and food debris.
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-

```

```

CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-
CC fructose + (1,6-alpha-D-glucosyl) (n+1).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes both
CC water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
CC forms of glucans.
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
CC -!- SIMILARITY: Contains 16 cell wall binding repeats.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC
CC EMBL; D90213; BAA14241.1; -; Genomic_DNA.
DR HSSP; P06653; 1GVM.
DR InterPro; IPR002479; Cell_wall_bd_put.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_I; 3.
DR Pfam; PF02324; Glyco_hydro_70; 1.
DR Dental caries; Glycosyltransferase; Repeat; Signal; Transferase.
FT SIGNAL 1 38
FT CHAIN 39 1592 Potential.
FT REPEAT 1093 1142 Glucosyltransferase-I.
FT REPEAT 1158 1207 1.
FT REPEAT 1222 1272 2.
FT REPEAT 1287 1337 3.
FT REPEAT 1402 1451 4.
FT REPEAT 1514 1563 5.
FT REPEAT 1577 1592 6.
FT REPEAT 1577 1592 7 (incomplete).
FT REGION 39 1044 Catalytic (approximate).
FT REGION 1093 1592 7 X tandem repeats.
FT REGION 1093 1592 Glucan-binding (approximate).
FT SEQUENCE 1592 AA; 176168 MW; BC0A66D079351ECF CRC64;

Query Match 100.0%; Score 121; DB 1; Length 1592;
Best Local Similarity 100.0%; Pred. No. 1.9e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNHVSIVEAWSNDTPYLHDD 21
Db 478 NNHVSIVEAWSNDTPYLHDD 498

RESULT 4
GTF1_STRDO
ID_GTF1_STRDO STANDARD; PRT; 1597 AA.
AC P11001;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranase)
DE (Sucrose 6-glucosyltransferase).
GN Name=gtf1;
OS Streptococcus downei (Streptococcus sobrinus).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OC NCBI_TaxID=1317;
RN [1]_TaxID=1317;
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC STRAIN=MFE28;
RX MEDLINE=87308014; PubMed=3040686;
RA Ferretti J.J., Gilpin M.L., Russell R.R.B.;
RA "Nucleotide sequence of a glucosyltransferase gene from Streptococcus
RT sobrinus MFE28.";
RL J. Bacteriol. 169:4271-4278(1987).
CC -!- FUNCTION: Production of extracellular glucans, that are thought to
CC play a key role in the development of the dental plaque because of
CC their ability to adhere to smooth surfaces and mediate the
CC aggregation of bacterial cells and food debris.
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-

```



```

FT VARIANT 399 399 strain MT8148).
FT VARIANT 474 512 N -> R (in strain MT4239).
FT VARIANT 512 474 K -> T (in strain MT4239).
FT VARIANT 519 519 F -> R (in strain MT8148).
FT VARIANT 701 701 T -> Y (in strain MT8148).
FT VARIANT 708 708 A -> V (in strain MT8148).
FT VARIANT 938 938 F -> L (in strain MT8148).
FT VARIANT 952 957 FOKPVE -> YGTPVA (in strain GS-5, strain MT4239 and strain MT4467).
FT VARIANT 963 964 SV -> NT (in strain GS-5, strain MT4239 and strain MT4467).
FT VARIANT 968 970 ADS -> VDG (in strain GS-5, strain MT4239 and strain MT4467).
FT VARIANT 1086 1086 A -> T (in strain MT4239).
FT VARIANT 1158 1158 S -> N (in strain MT4239).
FT VARIANT 1163 1163 H -> Y (in strain MT4251).
FT VARIANT 1168 1168 E -> K (in strain MT8148).
FT VARIANT 1182 1182 Y -> C (in strain MT8148).
FT VARIANT 1234 1234 A -> P (in strain MT4239).
FT VARIANT 1263 1263 R -> H (in strain GS-5 and strain MT4467).
FT VARIANT 1263 1263 R -> P (in strain MT8148).
FT VARIANT 1264 1264 Y -> H (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).
FT VARIANT 1272 1272 S -> G (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).
FT VARIANT 1329 1329 H -> Y (in strain GS-5 and strain MT4467).
FT VARIANT 1394 1394 Y -> H (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).
FT VARIANT 1402 1402 S -> G (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).
FT VARIANT 1459 1459 Y -> H (in strain MT8148).
FT CONFLICT 570 570 R -> A (in Ref. 1).
FT CONFLICT 800 817 ADQDVRAASTAPSTDGK -> LKMFALRLARPHQOMA (in Ref. 1).
FT CONFLICT 1310 1310 H -> L (in Ref. 1).
FT SEQUENCE 1476 AA; 165947 MW; 9C6E09F731B4CBFC CRC64;

Query Match 90.9%; Score 110; DB 1; Length 1476;
Best Local Similarity 85.7%; Pred. No. 9.3e-08;
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNHVSTVEAWSNDPTPLHDD 21
Db 482 NDHLSILEAWSNDPTPLHDD 502

RESULT 6
ID GTFC STRMU STANDARD; PRT; 1455 AA.
AC P13470; O69382; O69385; O69388; O69391; O69397; P05427;
DT 01-NOV-1988 (Rel. 09, Created)
DT 28-NOV-1988 (Rel. 41, Last sequence update)
DT 10-WAY-2005 (Rel. 47, Last annotation update)
DE Glucosyltransferase-SI precursor (EC 2.4.1.5) (GTF-SI)
DE (Dextranucrase) (Sucrose 6-glucosyltransferase).
GN Name=gTfC; OrddredLocusNames=SMU.1005;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=GS-5;
RX MEDLINE=89137980; PubMed=2976010; DOI=10.1016/0378-1119(88)90382-4;
RA Ueda S., Shiroza T., Kuramitsu H.K.;
RT "Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.";
RL Gene 69:101-109(1988).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MT4239 / Serotype c, MT4245 / Serotype e, MT4251 / Serotype f,

```

```

RC MT4467 / Serotype e, and MT8148 / Serotype c;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
RN [3]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX STRAIN=UA159 / ATCC 700610 / Serotype c;
RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;
RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Prineaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,
RA Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
RT pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
RN [4]
RP NUCLEOTIDE SEQUENCE OF 1-349.
RC STRAIN=GS-5;
RX MEDLINE=87308013; PubMed=3040685;
RA Shiroza T., Ueda S., Kuramitsu H.K.;
RT "Sequence analysis of the gtfB gene from Streptococcus mutans.";
RL J. Bacteriol. 169:4263-4270(1987).
CC -!- FUNCTION: Production of extracellular glucans, that are thought to
CC play a key role in the development of the dental plaque because of
CC their ability to adhere to smooth surfaces and mediate the
CC aggregation of bacterial cells and food debris.
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-
CC fructose + (1,6-alpha-D-glucosyl) (n+1).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes
CC water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
CC forms of glucans.
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
CC -!- SIMILARITY: Contains 5 cell wall binding repeats.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL; M22054; AAA88592.1; -; Genomic DNA.
CC EMBL; D88652; BAA26102.1; -; Genomic DNA.
CC EMBL; D88655; BAA26106.1; -; Genomic DNA.
CC EMBL; D88658; BAA26110.1; -; Genomic DNA.
CC EMBL; D88661; BAA26114.1; -; Genomic DNA.
CC EMBL; D89978; BAA26120.1; -; Genomic DNA.
CC EMBL; AF014940; AAN58706.1; -; Genomic DNA.
CC EMBL; M17361; AAA88589.1; -; Genomic DNA.
CC PIR; JTO345; JTO345.
CC HSSP; P06653; 1H9G.
CC InterPro; IPR002479; Cell wall_bd_put.
CC InterPro; IPR003318; Glyco_hydro_70.
CC Pfam; PF01473; CW_binding_1; 2.
CC Pfam; PF02324; Glyco_hydro_70; 1.
CC Complete proteome; Dental Caries; Glycosyltransferase; Repeat; Signal;
KW Transferase.
KW SIGNAL.
FT CHAIN 1 34 Glucosyltransferase-SI.
FT REPEAT 1126 1159 A repeat.
FT REPEAT 1169 1200 A repeat.
FT REPEAT 1227 1238 C repeat.
FT REPEAT 1253 1303 AC repeat.
FT REPEAT 1318 1330 A repeat (incomplete).
FT REGION 35 1050 Catalytic (approximate).
FT REGION 1126 1455 2.4 A, 1 C and 1 AC repeats.
FT REGION 1126 1455 Glucan-binding (approximate).
FT VARIANT 21 21 V -> I (in strain GS-5).
FT VARIANT 81 81 P -> L (in strain MT4239).

```


[1] NUCLEOTIDE SEQUENCE.
 RN STRAIN=GS-5; PubMed=2148600;
 RX MEDLINE=91100958; PubMed=2148600;
 RA Honda O., Kato C., Kuramitsu H.K.;
 RT "Nucleotide sequence of the Streptococcus mutans gtfD gene encoding
 the glucosyltransferase-S enzyme."; J. Gen. Microbiol. 136:2099-2105 (1990).
 RL [2]
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=MT4239 / Serotype c, MT4245 / Serotype e, MT4251 / Serotype f,
 RC MT4467 / Serotype e, and MT8148 / Serotype c;
 RX MEDLINE=98231643; PubMed=9570124;
 RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
 RA Kimura S., Hamada S.;
 RT "Molecular analyses of glucosyltransferase genes among strains of
 Streptococcus mutans"; FEMS Microbiol. Lett. 161:331-336 (1998).
 RL [3]
 RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RC STRAIN=UAI59 / ATCC 700610 / Serotype c;
 RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;
 RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
 RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
 RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,
 RA Ferretti J.J.;
 RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental
 pathogen."; Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439 (2002).
 RL CC
 CC -!- FUNCTION: Production of extracellular glucans, that are thought to
 play a key role in the development of the dental plaque because of
 their ability to adhere to smooth surfaces and mediate the
 aggregation of bacterial cells and food debris.
 CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-
 fructose + (1,6-alpha-D-glucosyl) (n+1).
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes
 water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
 forms of glucans.
 CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
 CC -!- SIMILARITY: Contains 6 cell wall binding repeats.
 CC
 CC This Swiss-Prot entry is copyrighted. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use as long as its content is in no way modified and this statement is not
 removed.
 CC
 CC EMBL; M29296; AAA26895.1; -; Genomic DNA.
 CC EMBL; D88653; BAA26103.1; -; Genomic DNA.
 CC EMBL; D88656; BAA26107.1; -; Genomic DNA.
 CC EMBL; D88659; BAA26111.1; -; Genomic DNA.
 CC EMBL; D88662; BAA26115.1; -; Genomic DNA.
 CC EMBL; D89979; BAA26121.1; -; Genomic DNA.
 CC EMBL; A0014932; AAN58619.1; -; Genomic DNA.
 CC HSSP; P06653; 1GVN.
 DR InterPro; IPR002479; Cell wall bd put.
 DR Pfam; PF01473; CW_binding_I; 4.
 DR Pfam; PF02324; Kglyco_hydro_70; 1.
 KW Complete proteome; Dental caries; Glycosyltransferase; Repeat; Signal;
 Transferase.
 FT SIGNAL 1 ? Potential.
 FT CHAIN ? 1462 Glucosyltransferase-S.
 FT REPEAT 1232 1295 1.
 FT REPEAT 1296 1359 2.
 FT REPEAT 1360 1423 3.
 FT REGION 1232 1423 3 X 63 AA approximate tandem repeats.
 FT VARIANT 10 10 Y -> H (in strain GS-5, strain MT4239,
 strain MT4245, strain MT4251, strain
 MT4467 and strain MT8148).
 FT VARIANT 19 19 I -> V (in strain GS-5, strain MT4239,

FT strain MT4245, strain MT4251, strain
 MT4467 and strain MT8148).
 FT K -> E (in strain MT4467).
 FT A -> S (in strain MT4239 and strain
 MT4245).
 FT A -> T (in strain MT4251 and strain
 MT8148).
 FT T -> I (in strain MT4239 and strain
 MT4245).
 FT A -> V (in strain MT4239, strain MT4245
 and strain MT8148).
 FT A -> S (in strain GS-5 and strain
 MT4467).
 FT A -> V (in strain MT4245).
 FT A -> T (in strain GS-5, strain MT4239,
 strain MT4245, strain MT4251, strain
 MT4467 and strain MT8148).
 FT V -> L (in strain MT4239).
 FT D -> N (in strain MT8148).
 FT E -> D (in strain MT4239, strain MT4245
 and strain MT4251).
 FT D -> N (in strain MT4239, strain MT4245
 and strain MT4251).
 FT Q -> H (in strain MT4245).
 FT D -> N (in strain MT4239 and strain
 MT4251).
 FT E -> K (in strain MT4239).
 FT V -> F (in strain MT4239).
 FT F -> L (in strain MT4239, strain MT4251
 and strain MT4467).
 FT KKQYQ -> EKEYTL (in strain MT4251).
 FT A -> S (in strain MT4239).
 FT TDQGEA -> ADKGNDS (in strain MT4251).
 FT TDQGS -> ADKGN (in strain MT4239 and
 strain MT4245).
 FT T -> A (in strain GS-5, strain MT4239,
 strain MT4245, strain MT4251, strain
 MT4467 and strain MT8148).
 FT D -> Y (in strain MT4251).
 FT E -> K (in strain MT4245 and strain
 MT4251).
 FT LG -> IR (in strain MT4251).
 FT G -> R (in strain MT4245).
 FT G -> R (in strain MT4239).
 FT H -> Q (in strain GS-5).
 FT S -> N (in strain MT4239).
 FT Y -> C (in strain MT4251 and strain
 MT4467).
 FT F -> L (in strain MT4467).
 FT Q -> P (in strain MT4245).
 FT K -> T (in strain MT4245).
 FT N -> D (in strain MT4245).
 FT G -> D (in strain GS-5 and strain
 MT4467).
 FT G -> R (in strain GS-5).
 FT R -> K (in strain MT4467).
 FT RYDNGSMVMVKKVVTLANGRRIGIDRWGIARYY -> VY
 R (in Ref. 1).
 SQ SEQUENCE 1462 AA; 163388 MW; CE4A279C4D708645 CRC64;
 Query Match 60.3%; Score 73; DB 1; Length 1462;
 Best Local Similarity 68.4%; Pred. No. 0.057;
 Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 QY 2 NHVSIVEAWSNDNDTPYLHD 20
 Db 497 NHLSLLEAWSNDNDPQYNKD 515
 ||:|||||
 ||:|||||
 RESULT 10
 Q8G9Q2 LEUME
 ID Q8G9Q2 LEUME PRELIMINARY; PRT; 2835 AA.
 AC Q8G9Q2;

RESULT	11
Q9LCH3_STROR	
ID	Q9LCH3_STROR PRELIMINARY; PRT; 1575 AA.
AC	Q9LCH3;
DT	01-OCT-2000 (TREMBLrel. 15, Created)
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT	01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE	Glucoyltransferase.
GN	Name=gtfr;
OS	Streptococcus oralis.
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OX	Streptococcus.
NCBI_TaxID	1303;
[1]	NUCLEOTIDE SEQUENCE.
RP	STRAIN=ATCC10557;
RX	MEDLINE=202311779; PubMed=10768934;
RX	DOI=10.1128/JAI.68.5.2475-2483.2000;
RA	Fujikawa T., Hoshino T., Oshima T., Sobue S., Hamada S.;
RT	"Purification, characterization, and molecular analysis of the gene encoding glucoyltransferase from Streptococcus oralis.";
RL	Infect. Immun. 68:2475-2483(2000).
EMB1	EMBL; AB025228; BAA95201.1; -; Genomic_DNA.
HSP	P06653; 1HCX.
GO	GO: 0016740; F:transferase activity; IEA.
GO	GO: 0009250; P:glucan biosynthesis; IEA.
InterPro	IPR002479; CW_binding_70.
InterPro	IPR003318; Glyco_hydro_70.
Pfam	PF01473; CW_binding_1; 4.
Pfam	Pf02324; Glyco_hydro_70; 1.
KW	Transferase.
SEQUENCE	1575 AA; 176792 MW; 772A26B4D7C2B543 CRC64;
SQ	
Query Match	55.4%; Score 67; DB 2; Length 1575;
Best Local Similarity	66.7%; Pred. NO. 0.54;

```

Qy      3  HVSIWEAWSNDNTPYLHD  20
         |:|:|:|:|:|:|:|
Db      551  HLSILEAWSNDPDYMKD  568

RESULT 13
Q9WXJ4_9STRE
ID      Q9WXJ4_9STRE PRELIMINARY;      PRT; 1338 AA.
AC      Q9WXJ4;
DT      01-NOV-1999 (trEMBLrel. 12, Created)
DT      01-NOV-1999 (trEMBLrel. 12, Last sequence update)
DT      01-JUN-2003 (trEMBLrel. 24, Last annotation update)
DE      GTF-S.stfs;
GN      Name=gtfs;
OS      Streptococcus criceti.
OG      Plasmid pAM1.
OC      Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC      Streptococcus.
OX      NCBI_TaxID=1333;
RN      [1]
RP      NUCLEOTIDE SEQUENCE.
RC      STRAIN=HS-6;
RA      Inoue M., Fukui K., Miyagi A.;
RL      Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.

```


FT	REPEAT	1150	1169	Cell wall binding 5.
FT	REPEAT	1170	1190	Cell wall binding 6.
FT	REPEAT	1225	1243	Cell wall binding 7.
FT	REPEAT	1289	1308	Cell wall binding 8.
FT	REPEAT	1309	1328	Cell wall binding 9.
FT	REPEAT	1331	1352	Cell wall binding 10.
FT	REGION	198	1061	Catalytic (approximate).
FT	SEQUENCE	1365 AA; 151591 MW; 167296BSA2E8C476 CRC64;		
Query Match 54.5%; Score 66; DB 1; Length 1365;				
Best Local Similarity 57.9%; Pred. No. 0.65;				
Matches 11; Conservative 4; Mismatches 4; Indels 0; Gaps 0;				
QY	2	NHVSIVEAWSNDTPTYLHD 20		
		: : : : : : :		
Db	469	DHLSILEANGSNDYVKD 487		
RESULT 15				
Q84CN4 LEUME				
ID	Q84CN4	LEUME PRELIMINARY;	PRT; 1330 AA.	
AC	Q84CN4;			
DT	01-JUN-2003	(TrEMBLrel. 24, Created)		
DT	01-JUN-2003	(TrEMBLrel. 24, Last sequence update)		
DT	01-MAR-2004	(TrEMBLrel. 26, Last annotation update)		
DE	Dextranucrase DsrR	(EC 2.4.1.5).		
GN	Name=dsrR;			
OS	Leuconostoc mesenteroides.			
OC	Bacteria; Firmicutes; Lactobacillales; Leuconostoc.			
OX	NCBI_TaxID=1245;			
RN	[1]			
RP	NUCLEOTIDE SEQUENCE.			
RC	STRAIN=NRRL B-1501;			
RA	Kim C.H., Moon J.O., Jang E.K.;			
RL	Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AY142210; AN38835.1; -; Genomic_DNA.			
DR	GO; GO:0047849; F:deoxytransucrase activity; IEA.			
DR	GO; GO:0016757; F:transucrase activity, transferring glycosyl. .; IEA.			
DR	GO; GO:0009250; P:glucan biosynthesis; IEA.			
DR	InterPro; IPR002479; CW binding.			
DR	InterPro; IPR003318; Glyco_hydro_70.			
DR	Pfam; PF01473; CW_binding_I.1.			
DR	Pfam; PF02324; Glyco_hydro_70; 1.			
KW	Glycosyltransferase; Transferase.			
SEQUENCE	1330 AA; 148863 MW; D945CBB36CF75797 CRC64;			
Query Match 52.9%; Score 64; DB 2; Length 1330;				
Best Local Similarity 55.0%; Pred. No. 1.3;				
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;				
QY	1	NNHVSIVEAWSNDTPTYLHD 20		
		: : : : : :		
Db	386	NOHLSILEDWNSNDPEYVKD 405		

Search completed: February 11, 2006, 19:39:14
Job time : 3.83362 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 18:56:56 ; Search time 1.99074 Seconds
(without alignments)
4414.224 Million cell updates/sec

Title: US-10-797-821-43

Perfect score: 105

Sequence: 1 VVIANNVDFVSWGITDFEM 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_21.*

- 1: Geneseqp1980s.*
- 2: Geneseqp1990s.*
- 3: Geneseqp2000s.*
- 4: Geneseqp2001s.*
- 5: Geneseqp2002s.*
- 6: Geneseqp2003as.*
- 7: Geneseqp2003bs.*
- 8: Geneseqp2004s.*
- 9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	105	100.0	20	ADX37286	Adx37286 Streptoco
2	105	100.0	1590	ADD93657	Add93657 Streptoco
3	105	100.0	1590	ADX37280	Adx37280 Streptoco
4	105	100.0	1592	AAR32925	Aar32925 Glucosylt
5	90	85.7	1017	AAU79285	Aau79285 Streptoco
6	90	85.7	1375	AAU98028	Aau98028 S. mutans
7	90	85.7	1375	AAU79288	Aau79288 Streptoco
8	90	85.7	1375	ADD93655	Add93655 Streptoco
9	90	85.7	1375	ADX37278	Adx37278 Streptoco
10	90	85.7	1475	AAU98031	Aau98031 S. mutans
11	90	85.7	1475	AAU98036	Aau98036 S. mutans
12	90	85.7	1475	AAU98037	Aau98037 S. mutans
13	90	85.7	1475	AAU98040	Aau98040 S. mutans
14	90	85.7	1475	AAU98035	Aau98035 S. mutans
15	90	85.7	1475	AAU98033	Aau98033 S. mutans
16	90	85.7	1475	AAU98034	Aau98034 S. mutans
17	90	85.7	1475	AAU98030	Aau98030 S. mutans
18	90	85.7	1475	AAU98039	Aau98039 S. mutans
19	90	85.7	1475	AAU98027	Aau98027 S. mutans
20	90	85.7	1475	AAU98032	Aau98032 S. mutans
21	90	85.7	1475	AAU98038	Aau98038 S. mutans
22	90	85.7	1475	ADD93654	Add93654 Streptoco
23	90	85.7	1475	ADX37277	Adx37277 Streptoco
24	90	85.7	1476	AAU79284	Aau79284 Streptoco

25	72	68.6	855	5	ABB98573	Abb98573 Dextran s
26	72	68.6	2835	5	ABB98574	Abb98574 Dextran s
27	72	68.6	2835	6	ABR55594	AbR55594 Amino aci
28	71	67.6	1518	7	ADD93660	Add93660 Streptoco
29	71	67.6	1518	9	ADX37283	Adx37283 Streptoco
30	69	65.7	1365	7	ADD93659	Add93659 Streptoco
31	69	65.7	1365	9	ADX37282	Adx37282 Streptoco
32	69	65.7	1499	7	ADC54806	Adc54806 Protein S
33	68	64.8	1527	5	AU80055	Au80055 Leuconost
34	68	64.8	1527	7	ADC54807	Adc54807 Leuconost
35	65	61.9	1430	5	AAU98043	Aau98043 S. mutans
36	65	61.9	1430	5	AAU98044	Aau98044 S. mutans
37	65	61.9	1430	5	AAU98045	Aau98045 S. mutans
38	65	61.9	1430	5	AAU98042	Aau98042 S. mutans
39	65	61.9	1430	5	AAU98029	Aau98029 S. mutans
40	65	61.9	1430	5	AAU98041	Aau98041 S. mutans
41	65	61.9	1430	7	ADD93656	Add93656 Streptoco
42	65	61.9	1430	9	ADX37279	Adx37279 Streptoco
43	65	61.9	2147	6	ABR63231	AbR63231 Glucansuc
44	63	60.0	2055	6	ABR63235	AbR63235 Glucansuc
45	56	53.3	1577	2	AAR91047	Aar91047 Alpha-D-g

ALIGNMENTS

RESULT 1
ADX37286
ID ADX37286 standard; peptide; 20 AA.
XX
AC ADX37286;
XX
DT 21-APR-2005 (first entry)
XX
DE Streptococcus sobrinus glucan binding protein B peptide #2.
XX
KW immunogenicity; immune stimulation; glucan binding protein-B;
KW microparticle; major histocompatibility complex; tooth disease.
XX
OS Streptococcus sobrinus.
XX
PN US2005031633-A1.
XX
PD 10-FEB-2005.
XX
PF 09-MAR-2004; 2004US-00797821.
XX
PR 13-APR-1998; 98US-0081550P.
PR 08-JAN-1999; 99US-0115142P.
PR 12-APR-1999; 99US-00290049.
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2003US-0402483P.
PR 07-MAR-2003; 2003US-00383930.
XX
(SMIT/) SMITH D J.
(TAUB/) TAUBMAN M A.
PI Smith DJ, Taubman MA;
XX WPI; 2005-151644/16.
DR
PT New composition comprising a fragment of a glucan binding protein-B (GbpB) that binds to MHC class II protein, and a biocompatible microparticle, useful for producing an antibody (claimed) for immunizing mammals against dental caries.
XX
PT Example 5; SEQ ID NO 43; 73pp; English.
XX
CC The invention relates to a composition comprising a fragment of a glucan binding protein-B (GbpB) and a biocompatible microparticle, where the fragment binds to a major histocompatibility complex (MHC) class II protein. The composition is useful for producing an antibody for immunizing mammals against dental caries. This sequence corresponds to a

```

CC Streptococcus sobrinus GbpB-derived peptide of the invention.
XX
SQ Sequence 20 AA;

  Query Match      100.0%; Score 105; DB 9; Length 20;
  Best Local Similarity 100.0%; Pred. No. 2.3e-09;
  Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVIANNVDFVSWGITDFEM 20
   |||||
Db 1 VVIANNVDFVSWGITDFEM 20

RESULT 2
ADD93657
ID ADD93657 standard; protein; 1590 AA.
XX
AC ADD93657;
XX
DT 29-JAN-2004 (first entry)
XX
DE Streptococcus sobrinus glucosyltransferase-I.
XX
KW Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.
XX
OS Streptococcus sobrinus.
XX
PN WO2003075845-A2.
XX
PD 18-SEP-2003.
XX
PF 07-MAR-2003; 2003WO-US006962.
XX
PR 07-MAR-2002; 2002US-0363209P.
XX
PR 08-AUG-2002; 2002US-0402483P.
XX
PA (FORS-) FORSYTH INST.
XX
PI Smith DJ, Taubman MA;
XX
PI WPI; 2003-845091/78.
XX
PT Composition useful as vaccines for dental caries comprises a fragment of
PT a glucan binding protein-B binding to a major histocompatibility complex
PT class II protein.
XX
PS Claim 16; Page 14; 49pp; English.
XX
CC The present sequence is the protein sequence of Streptococcus sobrinus
CC glucosyltransferase-I (GTF-I). Peptide fragments of GTF-I, especially
CC from the catalytic domain of the polypeptide, can be used in immunogenic
CC compositions and subunit vaccines for dental caries. These compositions
CC comprise a major histocompatibility complex (MHC) class II protein-
CC binding peptide from S. mutants glucan binding protein-B (GbpB)
CC covalently linked with a peptide fragment of a streptococcal
CC glucosyltransferase. The compositions are used in a claimed method of
CC eliciting production of an antibody in a mammal. Dieptopic or
CC multieptopic polypeptides can be prepared synthetically or by
CC recombinant DNA technology. Antibodies raised against MHC class II
CC binding fragments of GbpB can be used in passive immunisation.
XX
SQ Sequence 1590 AA;

  Query Match      100.0%; Score 105; DB 7; Length 1590;
  Best Local Similarity 100.0%; Pred. No. 4.1e-07;
  Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVIANNVDFVSWGITDFEM 20
   |||||
Db 847 VVIANNVDFVSWGITDFEM 866

RESULT 3

  Query Match      100.0%; Score 105; DB 7; Length 1590;
  Best Local Similarity 100.0%; Pred. No. 4.1e-07;
  Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVIANNVDFVSWGITDFEM 20
   |||||
Db 847 VVIANNVDFVSWGITDFEM 866

RESULT 4

  Query Match      100.0%; Score 105; DB 9; Length 1590;
  Best Local Similarity 100.0%; Pred. No. 4.1e-07;
  Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVIANNVDFVSWGITDFEM 20
   |||||
Db 847 VVIANNVDFVSWGITDFEM 866

RESULT 4
AAR32925
ID AAR32925 standard; protein; 1592 AA.
XX
AC AAR32925;
XX
DT 28-JUN-1993 (first entry)
XX
DE Glucosyltransferase I.
XX
KW GT-1; Streptococcus; dental; caries.
XX
OS Streptococcus sobrinus.
XX

```

PN JP05023188-A.
 XX 02-FEB-1993.
 XX 25-JUL-1991; 91JP-00186592.
 XX 25-JUL-1991; 91JP-00186592.
 XX (KATO/) KATO K.
 PA (FUKU/) FUKUI I.
 XX
 XX WPI; 1993-0794449/10.
 DR N-PSDB; AAQ37760.
 XX
 PT DNA sequence glucosyltransferase-I - comprises Streptococcus sobrinus
 PT DNA sequence with at least one nucleotide added or deleted.
 XX
 XX Claim 13; Page 15; 29pp; Japanese.
 XX
 CC The DNA sequence from Streptococcus sobrinus strain 6715 encodes
 CC glucosyltransferase-I (and mutants). The DNA was obtd. by treating S.
 CC sobrinus 6715 with mutanolysin, extracting the chromosomal DNA, partially
 CC digesting with Sau3AI and fractionating on agarose gel. The 3-5 kbp
 CC fragment was ligated into pUC18 and E. coli JM109 transformed with it. A
 CC GT-1 expressing clone was isolated and sequenced. The clone may be used
 CC in the development of a drug for dental caries
 XX
 XX Sequence 1592 AA;
 SQ
 Query Match 100.0%; Score 105; DB 2; Length 1592;
 Best Local Similarity 100.0%; Pred. No. 4.1e-07;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VVIANNVDFVSWGITDFEM 20
 |||||
 DB 847 VVIANNVDFVSWGITDFEM 866
 |||||
 RESULT 5
 AAU79285
 ID AAU79285 standard; protein; 1017 AA.
 XX
 AC AAU79285;
 XX
 XX 13-AUG-2002 (first entry)
 DT
 XX Streptococcus mutans monoclonal antibody-related protein #2.
 DE
 XX Antibody; dental caries; water insoluble glucan synthetase; anti-carries;
 KW glucosyl transferase-B; immunotherapy.
 XX Streptococcus mutans.
 OS
 XX JP2002114709-A.
 PN
 XX 16-APR-2002.
 PD
 XX 04-OCT-2000; 2000JP-00304889.
 PF
 XX 04-OCT-2000; 2000JP-00304889.
 PR
 XX (UYNI-) UNIV NIPPON.
 PA
 XX WPI; 2002-448885/48.
 DR
 XX Anti-carries agent composed of a monoclonal antibody against an inhibitory
 PT enzyme against water insoluble glucan synthetase of glucosyl transferase-
 PT B (GTF-B) of Streptococcus mutans.
 XX
 XX Claim 4; Page 17-19; 28pp; Japanese.
 PS
 XX The invention relates to a monoclonal antibody against dental caries and
 CC an anti-carries agent composed of a monoclonal antibody produced by

CC Streptococcus mutans, particularly mouse-hybridoma MHP126 (FERM P-17566)
 CC or mouse-hybridoma MHP136 (FERM P-17567), against an enzyme having
 CC inhibitive activity against water insoluble glucan synthetase of glucosyl
 CC transferase-B. The monoclonal antibody specifically inhibits water
 CC insoluble glucan synthetase of Streptococcus mutans produced glucosyl
 CC transferase-B and is used in the immunotherapy of dental caries. This
 CC sequence represents a Streptococcus mutans monoclonal antibody-related
 CC protein
 XX
 XX Sequence 1017 AA;
 SQ
 Query Match 85.7%; Score 90; DB 5; Length 1017;
 Best Local Similarity 80.0%; Pred. No. 5.5e-05;
 Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 VVIANNVDFVSWGITDFEM 20
 |||||
 DB 816 VVIANNVDFVSWGITDFEM 835
 |||||
 RESULT 6
 AAU98028
 ID AAU98028 standard; protein; 1375 AA.
 XX
 AC AAU98028;
 XX
 XX 27-AUG-2002 (first entry)
 DT
 XX S. mutans glucosyltransferase GTFC.
 DE
 XX Glucosyltransferase; GTFC; transgenic plant; paper sizing;
 KW coating composition; glucan; starch; latex; thermoplastic molecule;
 KW amyloplast; vacuole; paper manufacture.
 XX Streptococcus mutans.
 OS
 XX US2002031826-A1.
 PN
 XX 14-MAR-2002.
 PD
 XX 19-DEC-2000; 2000US-00740274.
 PF
 XX 07-JUN-1995; 95US-00478704.
 PR
 XX 07-JUN-1995; 95US-00482711.
 PR
 XX 07-JUN-1995; 95US-00485243.
 PR
 XX 16-JAN-1998; 98US-00007999.
 PR
 XX 16-JAN-1998; 98US-00088172.
 PR
 XX 20-JAN-1998; 98US-00009620.
 PR
 XX 11-DEC-1998; 98US-00210361.
 XX (NICH/) NICHOLS S E.
 PA
 XX Nichols SE;
 PI
 XX WPI; 2002-414332/44.
 DR
 XX N-PSDB; ABK52939.
 DR
 XX Glucosyltransferase B or D protein useful for producing a glucan useful
 PT as substitutes for and additions to modified starch and latexes in paper
 PT manufacture, comprises mutations in specific positions.
 XX
 XX Disclosure; Page 30-33; 44pp; English.
 PS
 XX The invention an isolated protein comprising a glucosyltransferase (GTF)
 CC B polypeptide having changes at position from I448V, D457N, D567E,
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,
 CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
 CC GTF mutant, an isolated polynucleotide which encodes PI or P2, or its
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the
 CC GTF mutant, an expression cassette comprising the polynucleotide operably
 CC linked to a promoter, a vector comprising the expression cassette, host

CC cell introduced with the vector, a transgenic plant comprising the
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
 CC coating composition comprising a glucan produced in a plant transformed
 CC with a gene encoding the mutant GFP, wild type or, starch, a latex,
 CC thermoplastic molecule or their combinations or glucan and starch where
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line
 CC deficient in starch biosynthesis, transformed with a gene encoding a
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
 CC comprising the glucan (paper sizing/coating agent). The vector is useful
 CC for producing a glucan in a plant. The method comprises transforming a
 CC plant cell with the vector, growing the plant cell under plant growing
 CC conditions to produce a regenerated plant and inducing expression of the
 CC polynucleotide for a time sufficient to produce the glucan in the
 CC regenerated plant, where the vector contains a transit sequence from
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.
 CC Glucans are useful as substitutes for and additions to modified starch
 CC and latexes in paper manufacture. Unlike prior art techniques, which
 CC require input materials that produce chemical effluents, paper
 CC manufacture utilising the glucan produced by GFP, which utilises
 CC biologically produced input materials, is more cost-effective and
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
 CC properties and impart gloss to the paper during coating step. The present
 CC sequence represents GFPc
 XX
 SQ Sequence 1375 AA;

Query Match 85.7%; Score 90; DB 5; Length 1375;

Best Local Similarity 80.0%; Pred. No. 7.8e-05;

Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVIANNVDKFSVSGITDFEM 20

||||| ||||| ||:|||||

Db 876 VVIANKVDKFAEWGVTDFEM 895

RESULT 7

AAU79288

ID AAU79288 standard; protein; 1375 AA.

AC AAU79288;

DT 13-AUG-2002 (first entry)

XX Streptococcus mutans monoclonal antibody-related protein #5.

XX Antibody; dental caries; water insoluble glucan synthetase; anti-carries;

KW glucosyl transferase-B; immunotherapy.

XX Streptococcus mutans.

OS Streptococcus mutans.

PN JP2002114709-A.

XX 16-APR-2002.

XX 04-OCT-2000; 2000JP-00304889.

XX 04-OCT-2000; 2000JP-00304889.

XX (UYNI-) UNIV NIPPON.

XX WPI; 2002-448885/48.

XX Anti-carries agent composed of a monoclonal antibody against an inhibitory
 PT enzyme against water insoluble glucan synthetase of glucosyl transferase-
 PT B (GTF-B) of Streptococcus mutans.

PS Disclosure; Page 22-25; 28pp; Japanese.

XX The invention relates to a monoclonal antibody against dental caries and
 CC an anti-carries agent composed of a monoclonal antibody produced by
 CC Streptococcus mutans, particularly mouse-hybridoma MHP126 (FERM P-17566)

CC or mouse-hybridoma MHP136 (FERM P-17567), against an enzyme having
 CC inhibitive activity against water insoluble glucan synthetase of glucosyl
 CC transferase-B. The monoclonal antibody specifically inhibits water
 CC insoluble glucan synthetase of Streptococcus mutans produced glucosyl
 CC transferase-B and is used in the immunotherapy of dental caries. This
 CC sequence represents a Streptococcus mutans monoclonal antibody-related
 CC protein
 XX
 SQ Sequence 1375 AA;

Query Match 85.7%; Score 90; DB 5; Length 1375;

Best Local Similarity 80.0%; Pred. No. 7.8e-05;

Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVIANNVDKFSVSGITDFEM 20

||||| ||||| ||:|||||

Db 876 VVIANKVDKFAEWGVTDFEM 895

RESULT 8

ADD93655

ID ADD93655 standard; protein; 1375 AA.

XX ADD93655;

XX 29-JAN-2004 (first entry)

XX Streptococcus mutans glucosyltransferase-C.

XX Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.

XX Streptococcus mutans.

XX WO2003075845-A2.

XX 18-SEP-2003.

XX 07-MAR-2003; 2003WO-US006962.

XX 07-MAR-2002; 2002US-0363209P.

XX 08-AUG-2002; 2002US-0402483P.

XX (FORS-) FORSYTH INST.

XX Smith DJ, Taubman MA;

XX WPI; 2003-845091/78.

XX Composition useful as vaccines for dental caries comprises a fragment of
 PT a glucan binding protein-B binding to a major histocompatibility complex
 PT class II protein.

XX Claim 16; Page 13; 49pp; English.

XX The present sequence is the protein sequence of Streptococcus mutans
 CC glucosyltransferase-C (GTF-C). Peptide fragments of GTF-C, especially
 CC from the catalytic domain of the polypeptide, can be used in immunogenic
 CC compositions and subunit vaccines for dental caries. These compositions
 CC comprise a major histocompatibility complex (MHC) class II protein-
 CC binding peptide from S. mutans glucan binding protein-B (GbpB)
 CC covalently linked with a peptide fragment of a streptococcal
 CC glucosyltransferase. The compositions are used in a claimed method of
 CC eliciting production of an antibody in a mammal. Dieptopic or
 CC multipitopic polypeptides can be prepared synthetically or by
 CC recombinant DNA technology. Antibodies raised against MHC class II
 CC binding fragments of GbpB can be used in passive immunisation.

XX Sequence 1375 AA;

Query Match 85.7%; Score 90; DB 7; Length 1375;

Best Local Similarity 80.0%; Pred. No. 7.8e-05;

Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 VVIANNVDKFSWGVTDFEM 20
 Db 876 VVIANNVDKFAEWGVTDFEM 895

RESULT 9

ADK37278
 ID ADK37278 standard; protein; 1375 AA.

XX AC ADK37278;
 XX 21-APR-2005 (first entry)
 XX Streptococcus mutant glucan binding protein B variant #7.
 DE immunogenicity; immune stimulation; glucan binding protein-B;
 KW microparticle; major histocompatibility complex; tooth disease.
 XX Streptococcus mutans.
 OS Streptococcus mutans.
 XX US2005031633-A1.
 PN 10-FEB-2005.
 PD 09-MAR-2004; 2004US-00797821.
 PF 13-APR-1998; 98US-0081550P.
 XX 08-JAN-1999; 99US-0115142P.
 PR 12-APR-1999; 99US-00290049.
 PR 07-MAR-2002; 2002US-0363209P.
 PR 08-AUG-2002; 2002US-0402483P.
 PR 07-MAR-2003; 2003US-00383930.
 XX (SMIT/) SMITH, J.
 PA (TAUB/) TAUBMAN M A.
 PI Smith DJ, Taubman MA;
 XX WPI; 2005-151644/16.

XX New composition comprising a fragment of a glucan binding protein-B
 PT (GbpB) that binds to MHC class II protein, and a biocompatible
 PT microparticle, useful for producing an antibody (claimed) for immunizing
 PT mammals against dental caries.
 XX Claim 7; SEQ ID NO 35; 73pp; English.
 PS The invention relates to a composition comprising a fragment of a glucan
 CC binding protein-B (GbpB) and a biocompatible microparticle, where the
 CC fragment binds to a major histocompatibility complex (MHC) class II
 CC protein. The composition is useful for producing an antibody for
 CC immunizing mammals against dental caries. This sequence corresponds to a
 CC Streptococcus mutans GbpB protein of the invention.
 XX Sequence 1375 AA;

Query Match 85.7%; Score 90; DB 9; Length 1375;
 Best Local Similarity 80.0%; Pred. No. 7.8e-05;
 Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 VVIANNVDKFSWGVTDFEM 20
 Db 876 VVIANNVDKFAEWGVTDFEM 895

RESULT 10

AAU98031
 ID AAU98031 standard; protein; 1475 AA.

XX AC AAU98031;
 XX 27-AUG-2002 (first entry)
 DT
 XX

DE S. mutans glucosyltransferase GTFB mutant D457N.
 XX Glucosyltransferase; GTFB; transgenic plant; paper sizing;
 KW coating composition; glucan; starch; latex; thermoplastic molecule;
 KW amyloplast; vacuole; paper manufacture; mutant; mutein.
 XX Streptococcus mutans.
 OS Synthetic.
 XX Key Location/Qualifiers
 FH Misc-difference 457 /note= "Wild-type Asp substituted by Asn"
 FT US2002031826-A1.
 XX 14-MAR-2002.
 XX 19-DEC-2000; 2000US-00740274.
 PF 07-JUN-1995; 95US-00478704.
 XX 07-JUN-1995; 95US-00482711.
 PR 07-JUN-1995; 95US-00485243.
 PR 16-JAN-1998; 98US-00007999.
 PR 16-JAN-1998; 98US-00008172.
 PR 20-JAN-1998; 98US-00009620.
 PR 11-DEC-1998; 98US-00210361.
 XX (NICH/) NICHOLS S B.
 PA Nichols SE;
 PI WPI; 2002-414332/44.

DR Glucosyltransferase B or D protein useful for producing a glucan useful
 PT as substitutes for and additions to modified starch and latexes in paper
 PT manufacture, comprises mutations in specific positions.
 XX Claim 36; Page; 44pp; English.
 PS The invention an isolated protein comprising a glucosyltransferase (GTF)
 CC B polypeptide having changes at position from 148V, D457N, D567T,
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T,
 CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the
 CC GTF mutant, an expression cassette comprising the polynucleotide operably
 CC linked to a promoter, a vector comprising the expression cassette, host
 CC cell introduced with the vector, a transgenic plant comprising the
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
 CC coating composition comprising a glucan produced in a plant transformed
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
 CC thermoplastic molecule or their combinations or glucan and starch where
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line
 CC deficient in starch biosynthesis, transformed with a gene encoding a
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
 CC comprising the glucan (paper sizing/coating agent). The vector is useful
 CC for producing a glucan in a plant. The method comprises transforming a
 CC plant cell with the vector, growing the plant cell under plant growing
 CC conditions to produce a regenerated plant and inducing expression of the
 CC polynucleotide for a time sufficient to produce the glucan in the
 CC regenerated plant, where the vector contains a transit sequence from
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.
 CC Glucans are useful as substitutes for and additions to modified starch
 CC and latexes in paper manufacture. Unlike prior art techniques, which
 CC require input materials that produce chemical effluents, paper
 CC manufacture utilising the glucan produced by GTF, which utilises
 CC biologically produced input materials, is more cost-effective and
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
 CC properties and impart gloss to the paper during coating step. The present

CC sequence represents a GTFB mutant of the invention. Note: The present
 CC sequence is not shown in the specification but was created by the indexer
 CC using the GTFB sequence appearing as AAU98027 and the information in
 CC claim 36

XX SQ Sequence 1475 AA;

Query Match 85.7%; Score 90; DB 5; Length 1475;
 Best Local Similarity 80.0%; Pred. No. 8.5e-05;
 Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVIANNVDKFSWGVTDFEM 20
 ||||| ||||| ||:|||||
 Db 849 VVIANKVDKFAEWGVTDFEM 868

RESULT 11

AAU98036 |
 ID AAU98036 standard; protein; 1475 AA.

XX AC AAU98036;

XX DT 27-AUG-2002 (first entry)

XX DE S. mutans glucosyltransferase GTFB mutant D567T/D571K.

XX KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;

XX KW coating composition; glucan; starch; latex; thermoplastic molecule;

XX KW amyloplast; vacuole; paper manufacture; mutant; mutein.

XX OS Streptococcus mutans.

XX OS Synthetic.

XX FH Key Location/Qualifiers

FT FT Misc-difference 567

FT FT /note= "Wild-type Asp substituted by Thr"

FT FT Misc-difference 571

FT FT /note= "Wild-type Asp substituted by Lys"

XX US2002031826-A1.

XX PN 14-MAR-2002.

XX PF 19-DEC-2000; 2000US-00740274.

XX PR 07-JUN-1995; 95US-00478704.

XX PR 07-JUN-1995; 95US-00482711.

XX PR 07-JUN-1995; 95US-00485243.

XX PR 16-JAN-1998; 98US-00007999.

XX PR 16-JAN-1998; 98US-00008172.

XX PR 20-JAN-1998; 98US-00009620.

XX PR 11-DEC-1998; 98US-00210361.

XX (NICH/) NICHOLS S E.

XX PA Nichols SE;

XX PI Nichols SE;

XX DR WPI; 2002-414332/44.

XX Glucosyltransferase B or D protein useful for producing a glucan useful
 PT as substitutes for and additions to modified starch and latexes in paper
 PT manufacture, comprises mutations in specific positions.

XX Claim 36; Page; 44pp; English.

XX The invention an isolated protein comprising a glucosyltransferase (GTF)
 CC B polypeptide having changes at position from I48V, D457N, D567T,
 CC K1014I, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,
 CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
 CC GTF mutant, an isolated polynucleotide which encodes p1 or p2, or its
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the

CC GTF mutant, an expression cassette comprising the polynucleotide operably
 CC linked to a promoter, a vector comprising the expression cassette, host
 CC cell introduced with the vector, a transgenic plant comprising the
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
 CC coating composition comprising a glucan produced in a plant transformed
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
 CC thermoplastic molecule or their combinations or glucan and starch where
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line
 CC deficient in starch biosynthesis, transformed with a gene encoding a
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
 CC comprising the glucan (paper sizing/coating agent). The vector is useful
 CC for producing a glucan in a plant. The method comprises transforming a
 CC plant cell with the vector, growing the plant cell under plant growing
 CC conditions to produce a regenerated plant and inducing expression of the
 CC polynucleotide for a time sufficient to produce the glucan in the
 CC regenerated plant, where the vector contains a transit sequence from
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.
 CC Glucans are useful as substitutes for and additions to modified starch
 CC and latexes in paper manufacture. Unlike prior art techniques, which
 CC require input materials that produce chemical effluents, paper
 CC manufacture utilising the glucan produced by GTF, which utilises
 CC biologically produced input materials, is more cost-effective and
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
 CC properties and impart gloss to the paper during coating step. The present
 CC sequence represents a GTFB mutant of the invention. Note: The present
 CC sequence is not shown in the specification but was created by the indexer
 CC using the GTFB sequence appearing as AAU98027 and the information in
 CC claim 36

XX SQ Sequence 1475 AA;

Query Match 85.7%; Score 90; DB 5; Length 1475;

Best Local Similarity 80.0%; Pred. No. 8.5e-05;

Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVIANNVDKFSWGVTDFEM 20
 ||||| ||||| ||:|||||
 Db 849 VVIANKVDKFAEWGVTDFEM 868

RESULT 12

AAU98037

ID AAU98037 standard; protein; 1475 AA.

XX AC AAU98037;

XX DT 27-AUG-2002 (first entry)

XX DE S. mutans glucosyltransferase GTFB mutant D567T/D571K/K1014T.

XX KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;
 KW coating composition; glucan; starch; latex; thermoplastic molecule;
 KW amyloplast; vacuole; paper manufacture; mutant; mutein.

XX OS Streptococcus mutans.

XX OS Synthetic.

XX FH Key Location/Qualifiers

FT FT Misc-difference 567 /note= "Wild-type Asp substituted by Thr"

FT FT Misc-difference 571 /note= "Wild-type Asp substituted by Lys"

FT FT Misc-difference 1014 /note= "Wild-type Lys substituted by Thr"

XX US2002031826-A1.

XX PN 14-MAR-2002.

XX PD 19-DEC-2000; 2000US-00740274.

PR 07-JUN-1995; 95US-00478704.
 PR 07-JUN-1995; 95US-00482711.
 PR 07-JUN-1995; 95US-00485243.
 PR 16-JAN-1998; 98US-00007999.
 PR 16-JAN-1998; 98US-00008172.
 PR 20-JAN-1998; 98US-00009620.
 PR 11-DEC-1998; 98US-00210361.
 XX
 PA (NICH/) NICHOLS S E.
 XX
 PI Nichols SE;
 XX
 DR WPI; 2002-414332/44.
 XX
 PT Glucosyltransferase B or D protein useful for producing a glucan useful
 PT as substitutes for and additions to modified starch and latexes in paper
 PT manufacture, comprises mutations in specific positions.
 XX
 PS Claim 36; Page; 44pp; English.
 XX
 CC The invention an isolated protein comprising a glucosyltransferase (GTF)
 CC B polypeptide having changes at position from I448V, D457N, D567T,
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/D571K/K1014T,
 CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D, or a
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the
 CC GTF mutant, an expression cassette comprising the polynucleotide operably
 CC linked to a promoter, a vector comprising the expression cassette, host
 CC cell introduced with the vector, a transgenic plant comprising the
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
 CC coating composition comprising a glucan produced in a plant transformed
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
 CC thermoplastic molecule or their combinations or glucan and starch where
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line
 CC deficient in starch biosynthesis, transformed with a gene encoding a
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
 CC comprising the glucan (paper sizing/coating agent). The vector is useful
 CC for producing a glucan in a plant. The method comprises transforming a
 CC plant cell with the vector, growing the plant cell under plant growing
 CC conditions to produce a regenerated plant and inducing expression of the
 CC polynucleotide for a time sufficient to produce the glucan in the
 CC regenerated plant, where the vector contains a transit sequence from
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.
 CC Glucans are useful as substitutes for and additions to modified starch
 CC and latexes in paper manufacture. Unlike prior art techniques, which
 CC require input materials that produce chemical effluents, paper
 CC manufacture utilising the glucan produced by GTF, which utilises
 CC biologically produced input materials, is more cost-effective and
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
 CC properties and impart gloss to the paper during coating step. The present
 CC sequence represents a GTFB mutant of the invention. Note: The present
 CC sequence is not shown in the specification but was created by the indexer
 CC using the GTFB sequence appearing as AAU98027 and the information in
 CC claim 36
 XX
 SQ Sequence 1475 AA;
 Best Match 85.7%; Score 90; DB 5; Length 1475;
 Best Local Similarity 80.0%; Pred. No. 8.5e-05;
 Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 VVIANNVDFKSWGVTDFEM 20
 ||||| ||||| |||||
 Db 849 VVIANNVDFKSWGVTDFEM 868
 RESULT 13
 AAU98040
 ID AAU98040 standard; protein; 1475 AA.

XX
 AC AAU98040;
 DT 27-AUG-2002 (first entry)
 XX
 DE S. mutans glucosyltransferase GTFB mutant K779Q.
 XX
 KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;
 KW coating composition; glucan; starch; latex; thermoplastic molecule;
 KW amyloplast; vacuole; paper manufacture; mutant; mutein.
 XX
 OS Streptococcus mutans.
 OS Synthetic.
 XX
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 779 /note= "Wild-type Lys substituted by Gln"
 FT
 XX US2002031826-A1.
 XX
 PD 14-MAR-2002.
 XX
 PF 19-DEC-2000; 2000US-00740274.
 XX
 PR 07-JUN-1995; 95US-00478704.
 PR 07-JUN-1995; 95US-00482711.
 PR 07-JUN-1995; 95US-00485243.
 PR 16-JAN-1998; 98US-00007999.
 PR 16-JAN-1998; 98US-00008172.
 PR 20-JAN-1998; 98US-00009620.
 PR 11-DEC-1998; 98US-00210361.
 XX
 PA (NICH/) NICHOLS S E.
 XX
 PI Nichols SE;
 XX
 DR WPI; 2002-414332/44.
 XX
 PT Glucosyltransferase B or D protein useful for producing a glucan useful
 PT as substitutes for and additions to modified starch and latexes in paper
 PT manufacture, comprises mutations in specific positions.
 XX
 PS Claim 36; Page; 44pp; English.
 XX
 CC The invention an isolated protein comprising a glucosyltransferase (GTF)
 CC B polypeptide having changes at position from I448V, D457N, D567T,
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/D571K/K1014T,
 CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D, or a
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the
 CC GTF mutant, an expression cassette comprising the polynucleotide operably
 CC linked to a promoter, a vector comprising the expression cassette, host
 CC cell introduced with the vector, a transgenic plant comprising the
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
 CC coating composition comprising a glucan produced in a plant transformed
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
 CC thermoplastic molecule or their combinations or glucan and starch where
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line
 CC deficient in starch biosynthesis, transformed with a gene encoding a
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
 CC comprising the glucan (paper sizing/coating agent). The vector is useful
 CC for producing a glucan in a plant. The method comprises transforming a
 CC plant cell with the vector, growing the plant cell under plant growing
 CC conditions to produce a regenerated plant and inducing expression of the
 CC polynucleotide for a time sufficient to produce the glucan in the
 CC regenerated plant, where the vector contains a transit sequence from
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.
 CC Glucans are useful as substitutes for and additions to modified starch
 CC and latexes in paper manufacture. Unlike prior art techniques, which
 CC require input materials that produce chemical effluents, paper
 CC manufacture utilising the glucan produced by GTF, which utilises
 CC biologically produced input materials, is more cost-effective and
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
 CC properties and impart gloss to the paper during coating step. The present
 CC sequence represents a GTFB mutant of the invention. Note: The present
 CC sequence is not shown in the specification but was created by the indexer
 CC using the GTFB sequence appearing as AAU98027 and the information in
 CC claim 36
 XX
 SQ Sequence 1475 AA;
 Best Match 85.7%; Score 90; DB 5; Length 1475;
 Best Local Similarity 80.0%; Pred. No. 8.5e-05;
 Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 VVIANNVDFKSWGVTDFEM 20
 ||||| ||||| |||||
 Db 849 VVIANNVDFKSWGVTDFEM 868
 RESULT 13
 AAU98040
 ID AAU98040 standard; protein; 1475 AA.

CC require input materials that produce chemical effluents, paper
 CC manufacture utilising the glucan produced by GTF, which utilises
 CC biologically produced input materials, is more cost-effective and
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
 CC properties and impart gloss to the paper during coating step. The present
 CC sequence represents a GTFB mutant of the invention. Note: The present
 CC sequence is not shown in the specification but was created by the indexer
 CC using the GTFB sequence appearing as AAU98027 and the information in
 CC claim 36
 CC
 XX Sequence 1475 AA;
 SQ
 Query Match 85.7%; Score 90; DB 5; Length 1475;
 Best Local Similarity 80.0%; Pred. No. 8.5e-05;
 Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 VVIANNVDFKFSWGVTDFEM 20
 DB 849 VVIANNVDFKFAEWGVTDFEM 868
 RESULT 14
 AAU98035
 ID AAU98035 standard; protein; 1475 AA.
 XX AAU98035;
 AC
 XX 27-AUG-2002 (first entry)
 DT
 DE S. mutans glucosyltransferase GTFB mutant D457N/D571K.
 XX
 XX Glucosyltransferase; GTFB; transgenic plant; paper sizing;
 KW coating composition; glucan; starch; latex; thermoplastic molecule;
 KW amyloplast; vacuole; paper manufacture; mutant; mutein.
 XX
 OS Streptococcus mutans.
 OS Synthetic.
 XX
 XX Key Location/Qualifiers
 FH Misc-difference 457 /note= "Wild-type Asp substituted by Asn"
 FT Misc-difference 571 /note= "Wild-type Asp substituted by Lys"
 FT
 XX US2002031826-A1.
 PN
 XX 14-MAR-2002.
 XX
 XX 19-DEC-2000; 2000US-00740274.
 XX
 PR 07-JUN-1995; 95US-00478704.
 PR 07-JUN-1995; 95US-00482711.
 PR 07-JUN-1995; 95US-00485243.
 PR 16-JAN-1998; 98US-00007999.
 PR 16-JAN-1998; 98US-00008172.
 PR 20-JAN-1998; 98US-00009620.
 PR 11-DEC-1998; 98US-00210361.
 XX
 XX (NICHOLS) NICHOLS S E.
 PA
 XX Nichols SE;
 PI
 XX WPI; 2002-414332/44.
 XX
 XX Glucosyltransferase B or D protein useful for producing a glucan useful
 PT as substitutes for and additions to modified starch and latexes in paper
 PT manufacture, comprises mutations in specific positions.
 PT
 XX Claim 36; Page; 44pp; English.
 PS
 XX The invention an isolated protein comprising a glucosyltransferase (GTF)
 CC B polypeptide having changes at position from I448V, D457N, D567T,
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/K1014T,
 CC

CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the
 CC GTF mutant, an expression cassette comprising the polynucleotide operably
 CC linked to a promoter, a vector comprising the expression cassette, host
 CC cell introduced with the vector, a transgenic plant comprising the
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
 CC coating composition comprising a glucan produced in a plant transformed
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
 CC thermoplastic molecule or their combinations or glucan and starch where
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line
 CC deficient in starch biosynthesis, transformed with a gene encoding a
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
 CC comprising the glucan (paper sizing/coating agent). The vector is useful
 CC for producing a glucan in a plant. The method comprises transforming a
 CC plant cell with the vector, growing the plant cell under plant growing
 CC conditions to produce a regenerated plant and inducing expression of the
 CC polynucleotide for a time sufficient to produce the glucan in the
 CC regenerated plant, where the vector contains a transit sequence from
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.
 CC Glucans are useful as substitutes for and additions to modified starch
 CC and latexes in paper manufacture. Unlike prior art techniques, which
 CC require input materials that produce chemical effluents, paper
 CC manufacture utilising the glucan produced by GTF, which utilises
 CC biologically produced input materials, is more cost-effective and
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
 CC properties and impart gloss to the paper during coating step. The present
 CC sequence represents a GTFB mutant of the invention. Note: The present
 CC sequence is not shown in the specification but was created by the indexer
 CC using the GTFB sequence appearing as AAU98027 and the information in
 CC claim 36
 CC
 XX Sequence 1475 AA;
 SQ

Query Match 85.7%; Score 90; DB 5; Length 1475;
 Best Local Similarity 80.0%; Pred. No. 8.5e-05;
 Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVIANNVDFKFSWGVTDFEM 20
 DB 849 VVIANNVDFKFAEWGVTDFEM 868

RESULT 15

AAU98033
 ID AAU98033 standard; protein; 1475 AA.

XX AC AAU98033;

XX DT 27-AUG-2002 (first entry)

XX DE S. mutans glucosyltransferase GTFB mutant K1014T.

KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;
 KW coating composition; glucan; starch; latex; thermoplastic molecule;
 KW amyloplast; vacuole; paper manufacture; mutant; mutein.

XX OS Streptococcus mutans.

XX OS Synthetic.

XX FH Key Location/Qualifiers

FT Misc-difference 1014 /note= "Wild-type Lys substituted by Thr"

XX FT

XX PN US2002031826-A1.

XX PD 14-MAR-2002.

XX XX 19-DEC-2000; 2000US-00740274.

PF

```

XX 07-JUN-1995; 95US-00478704.
PR 07-JUN-1995; 95US-00482711.
PR 07-JUN-1995; 95US-00485243.
PR 16-JAN-1998; 98US-00007999.
PR 16-JAN-1998; 98US-00008172.
PR 20-JAN-1998; 98US-00009620.
PR 11-DEC-1998; 98US-00210361.
XX
XX (NICH/) NICHOLS S E.
XX
XX Nichols SE;
XX
XX WFI; 2002-414332/44.
XX
XX Glucosyltransferase B or D protein useful for producing a glucan useful
PT as substitutes for and additions to modified starch and latexes in paper
PT manufacture, comprises mutations in specific positions.
XX
XX Claim 36; Page; 4pp; English.
XX
XX The invention an isolated protein comprising a glucosyltransferase (GTF)
CC B polypeptide having changes at position from I448V, D457N, D567T,
CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/K1014T,
CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
CC complementary polynucleotide, a ribonucleic acid sequence encoding the
CC GTF mutant, an expression cassette comprising the polynucleotide operably
CC linked to a promoter, a vector comprising the expression cassette, host
CC cell introduced with the vector, a transgenic plant, a paper sizing and/or
CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
CC coating composition comprising a glucan produced in a plant transformed
CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
CC thermoplastic molecule or their combinations or glucan and starch where
CC the glucan is produced in the amyloplast and/or vacuole or a maize line
CC deficient in starch biosynthesis, transformed with a gene encoding a
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
CC comprising the glucan (paper sizing/coating agent). The vector is useful
CC for producing a glucan in a plant. The method comprises transforming a
CC plant cell with the vector, growing the plant cell under plant growing
CC conditions to produce a regenerated plant and inducing expression of the
CC polynucleotide, for a time sufficient to produce the glucan in the
CC regenerated plant, where the vector contains a transit sequence from
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
CC is produced in the amyloplast of potato or the vacuole of sugar beet.
CC Glucans are useful as substitutes for and additions to modified starch
CC and latexes in paper manufacture. Unlike prior art techniques, which
CC require input materials that produce chemical effluents, paper
CC manufacture utilising the glucan produced by GTF, which utilises
CC biologically produced input materials, is more cost-effective and
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
CC properties and impart gloss to the paper during coating step. The present
CC sequence represents a GTFB mutant of the invention. Note: The present
CC sequence is not shown in the specification but was created by the indexer
CC using the GTFB sequence appearing as AAU98027 and the information in
CC claim 36
XX
XX Sequence 1475 AA;
SQ
Query Match 85.7%; Score 90; DB 5; Length 1475;
Best Local Similarity 80.0%; Pred. No. 8.5e-05;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 VVIANNVDKFSVSGITDFEM 20
Db 849 VVIANNVDKFAEWSGVTDFEM 868

```

Search completed: February 11, 2006, 19:15:28
 Job time : 1.99074 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 19:39:30 ; Search time 0.437886 Seconds
(without alignments)
3776.130 Million cell updates/sec

Title: US-10-797-821-43

Perfect score: 105
Sequence: 1 VVIANNVDKFSWGITDFEM 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/1/iaa/5_COMB.pep.*
2: /cgn2_6/prodata/1/iaa/6_COMB.pep.*
3: /cgn2_6/prodata/1/iaa/H_COMB.pep.*
4: /cgn2_6/prodata/1/iaa/PCTUS_COMB.pep.*
5: /cgn2_6/prodata/1/iaa/RS_COMB.pep.*
6: /cgn2_6/prodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	90	85.7	1375	2	US-09-210-361-4
2	90	85.7	1375	2	US-09-740-274-4
3	90	85.7	1475	2	US-09-007-999-2
4	90	85.7	1475	2	US-09-210-361-2
5	90	85.7	1475	2	US-09-740-274-2
6	68	64.8	522	2	US-09-995-749A-11
7	68	64.8	522	2	US-09-604-957-5
8	65	61.9	545	2	US-09-604-957-4
9	65	61.9	545	2	US-09-995-749A-10
10	65	61.9	1430	2	US-09-008-172-2
11	65	61.9	1430	2	US-09-210-361-6
12	65	61.9	1430	2	US-09-740-274-6
13	56	53.3	1577	1	US-08-793-824-2
14	53	50.5	535	2	US-09-604-957-7
15	53	50.5	535	2	US-09-995-749A-13
16	53	50.5	1278	2	US-09-604-957-3
17	53	50.5	1781	2	US-09-995-749A-2
18	49	46.7	659	2	US-09-583-110-4043
19	49	46.7	661	2	US-09-107-433-3700
20	46	43.8	402	2	US-09-134-001C-4138
21	45.5	43.3	399	1	US-08-282-197C-54
22	45.5	43.3	837	2	US-09-390-234-14
23	45.5	43.3	837	2	US-09-603-311-14
24	45	42.9	584	2	US-09-604-957-6
25	45	42.9	584	2	US-09-995-749A-12
26	45	42.9	2057	2	US-09-499-203-2
27	44	41.9	810	1	US-08-785-241-7

28	43	41.0	102	1	US-08-480-473B-47	Sequence 47, Appl
29	43	41.0	102	2	US-08-915-213-47	Sequence 47, Appl
30	43	41.0	102	2	US-09-235-217-47	Sequence 47, Appl
31	43	41.0	202	2	US-08-949-155-2	Sequence 2, Appli
32	43	41.0	202	2	US-09-819-964-2	Sequence 2, Appli
33	43	41.0	212	2	US-09-328-352-6684	Sequence 6684, Ap
34	43	41.0	245	2	US-09-438-833-3	Sequence 3, Appli
35	43	41.0	330	2	US-09-438-833-4	Sequence 4, Appli
36	43	41.0	373	1	US-08-480-473B-3	Sequence 3, Appli
37	43	41.0	373	2	US-08-915-213-3	Sequence 3, Appli
38	43	41.0	373	2	US-09-235-217-3	Sequence 3, Appli
39	43	41.0	373	4	PCT-US96-10251-3	Sequence 3, Appli
40	43	41.0	613	2	US-09-438-833-6	Sequence 6, Appli
41	43	41.0	623	2	US-09-967-388-2	Sequence 2, Appli
42	43	41.0	652	2	US-09-438-833-5	Sequence 5, Appli
43	43	41.0	756	2	US-09-438-833-11	Sequence 11, Appl
44	43	41.0	805	1	US-08-480-473B-4	Sequence 4, Appli
45	43	41.0	805	2	US-08-915-213-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-210-361-4
; Sequence 4, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; EARLIER FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4

; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-210-361-4

Query Match 85.7%; Score 90; DB 2; Length 1375;
Best Local Similarity 80.0%; Pred. No. 2e-05;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVIANNVDKFSWGITDFEM 20
DB 876 VVIANNVDKFAEWGVTDFEM 895

RESULT 2
US-09-740-274-4
; Sequence 4, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19

```
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-4
```

```
Query Match 85.7%; Score 90; DB 2; Length 1375;
Best Local Similarity 80.0%; Pred. No. 2e-05;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 VVIANNVDKFSWGVTDFEM 20
||||| ||||| |||||
Db 876 VVIANNVDKFAEWGVTDFEM 895
```

```
RESULT 3
US-09-007-999-2
; Sequence 2, Application US/09007999
; Patent No. 6087559
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0356D
; CURRENT APPLICATION NUMBER: US/09/007,999
; CURRENT FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-007-999-2
```

```
Query Match 85.7%; Score 90; DB 2; Length 1475;
Best Local Similarity 80.0%; Pred. No. 2.1e-05;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 VVIANNVDKFSWGVTDFEM 20
||||| ||||| |||||
Db 849 VVIANNVDKFAEWGVTDFEM 868
```

```
RESULT 4
US-09-210-361-2
; Sequence 2, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
```

```
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-210-361-2
```

```
Query Match 85.7%; Score 90; DB 2; Length 1475;
Best Local Similarity 80.0%; Pred. No. 2.1e-05;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 VVIANNVDKFSWGVTDFEM 20
||||| ||||| |||||
Db 849 VVIANNVDKFAEWGVTDFEM 868
```

```
RESULT 5
US-09-740-274-2
; Sequence 2, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-740-274-2
```

```
Query Match 85.7%; Score 90; DB 2; Length 1475;
Best Local Similarity 80.0%; Pred. No. 2.1e-05;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 VVIANNVDKFSWGVTDFEM 20
||||| ||||| |||||
Db 849 VVIANNVDKFAEWGVTDFEM 868
```

```
RESULT 6
US-09-995-749A-11
; Sequence 11, Application US/09995749A
; Patent No. 6867026
```

RESULT 8
US-09-604-957-4
; Sequence 4, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
FILE REFERENCE: BO 43388

RESULT 10
US-09-008-172-2
Sequence 2, Application US/09008172
Patent No. 6127602
GENERAL INFORMATION:
APPLICANT: Nichols, Scott E.
TITLE OF INVENTION: Substitutes for Modified Starch and
TITLE OF INVENTION: Latexes in Paper Manufacture
FILE REFERENCE: 0358D
CURRENT APPLICATION NUMBER: US/09/008,172
CURRENT FILING DATE: 1998-01-16
EARLIER APPLICATION NUMBER: 08/482,711
EARLIER FILING DATE: 1995-06-07
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 2
LENGTH: 1430
TYPE: PRN

```
; ORGANISM: Streptococcus mutans
US-09-008-172-2

Query Match          61.9%; Score 65; DB 2; Length 1430;
Best Local Similarity 66.7%; Pred. No. 0.17;
Matches 12; Conservative 1; Mismatches 5; Indels 5; Gaps 0;

QY 3 IANNVDKFSWGTTDFEM 20
   ||||| ||||| |||||
Db 883 IAQNVQLFKSWGVTSEFEM 900

RESULT 11
US-09-210-361-6
; Sequence 6, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-210-361-6

Query Match          61.9%; Score 65; DB 2; Length 1430;
Best Local Similarity 66.7%; Pred. No. 0.17;
Matches 12; Conservative 1; Mismatches 5; Indels 5; Gaps 0;

QY 3 IANNVDKFSWGTTDFEM 20
   ||||| ||||| |||||
Db 883 IAQNVQLFKSWGVTSEFEM 900

RESULT 12
US-09-740-274-6
; Sequence 6, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172

Query Match          53.3%; Score 56; DB 1; Length 1577;
Best Local Similarity 61.1%; Pred. No. 4.7;
Matches 11; Conservative 0; Mismatches 7; Indels 7; Gaps 0;

QY 2 VIANNVDKFSWGTTDFE 19
   ||||| ||||| |||||
Db 958 VIAQNAKLFKEWGTSEF 975

; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-6

Query Match          61.9%; Score 65; DB 2; Length 1430;
Best Local Similarity 66.7%; Pred. No. 0.17;
Matches 12; Conservative 1; Mismatches 5; Indels 5; Gaps 0;

QY 3 IANNVDKFSWGTTDFEM 20
   ||||| ||||| |||||
Db 883 IAQNVQLFKSWGVTSEFEM 900

RESULT 13
US-08-793-824-2
; Sequence 2, Application US/08793824
; Patent No. 5981838
; GENERAL INFORMATION:
; APPLICANT: Simpson, Christine Lynn
; APPLICANT: Giffard, Philip Morrison
; APPLICANT: Jacques, Nicholas Anthony
; TITLE OF INVENTION: Genetic Manipulation of Plants to
; TITLE OF INVENTION: Increase Stored Carbohydrates
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Griffith Hack & Co
; STREET: Level 8, 168 Walker Street
; CITY: No. 5981838th Sydney
; STATE: New South Wales
; COUNTRY: Australia
; ZIP: 2060
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,824
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PM7643
; FILING DATE: 24-AUG-1994
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 61 2 9957 5944
; TELEFAX: 61 2 957 6288
; TELEX: 26547
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1577 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus salivarius
US-08-793-824-2

Query Match          53.3%; Score 56; DB 1; Length 1577;
Best Local Similarity 61.1%; Pred. No. 4.7;
Matches 11; Conservative 0; Mismatches 7; Indels 7; Gaps 0;

QY 2 VIANNVDKFSWGTTDFE 19
   ||||| ||||| |||||
Db 958 VIAQNAKLFKEWGTSEF 975
```

RESULT 14
US-09-604-957-7
; Sequence 7, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 535
; TYPE: PRT
; ORGANISM: Lactobacillus reuteri
US-09-604-957-7

Query Match 50.5%; Score 53; DB 2; Length 535;
Best Local Similarity 60.0%; Pred. No. 4;
Matches 12; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 VVIANNVDFKFSWGITDFEM 20
Db 451 VRIAQNADFFASLGFTSFEM 470

RESULT 15
US-09-995-749A-13
; Sequence 13, Application US/09995749A
; Patent No. 6867026
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: B043388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 535
; TYPE: PRT
; ORGANISM: Lactobacillus reuteri
US-09-995-749A-13

Query Match 50.5%; Score 53; DB 2; Length 535;
Best Local Similarity 60.0%; Pred. No. 4;
Matches 12; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 VVIANNVDFKFSWGITDFEM 20
Db 451 VRIAQNADFFASLGFTSFEM 470

Search completed: February 11, 2006, 19:46:20
Job time : 1.43789 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 20:39:23 ; Search time 1.79012 Seconds
(without alignments)
4668.162 Million cell updates/sec

Title: US-10-797-821-43

Perfect score: 105
Sequence: 1 VVIANNVDFVSWGITDFEM 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA_Main:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	105	100.0	20	5	US-10-797-821-43
2	105	100.0	1590	4	US-10-383-930-37
3	105	100.0	1590	5	US-10-797-821-37
4	90	85.7	1375	3	US-09-740-274-4
5	90	85.7	1375	4	US-10-383-930-35
6	90	85.7	1375	5	US-10-797-821-35
7	90	85.7	1475	3	US-09-740-274-2
8	90	85.7	1475	4	US-10-383-930-34
9	90	85.7	1475	5	US-10-797-821-34
10	71	67.6	1518	4	US-10-383-930-40
11	71	67.6	1518	5	US-10-797-821-40
12	69	65.7	1365	4	US-10-383-930-39
13	69	65.7	1365	5	US-10-797-821-39
14	68	64.8	522	3	US-09-995-749A-11
15	65	61.9	545	3	US-09-995-749A-10
16	65	61.9	1430	3	US-09-740-274-6
17	65	61.9	1430	4	US-10-383-930-36
18	65	61.9	1430	5	US-10-797-821-36
19	65	61.9	1777	5	US-10-484-218-12
20	63	60.0	1595	5	US-10-484-218-20
21	54	51.4	1006	5	US-10-484-218-22
22	53	50.5	535	3	US-09-995-749A-13
23	53	50.5	1781	3	US-09-995-749A-2
24	50	47.6	1554	4	US-10-383-930-38
25	50	47.6	1554	5	US-10-797-821-38
26	50	47.6	1771	5	US-10-484-218-14
27	49	46.7	493	5	US-10-472-928-444

28	49	46.7	661	5	US-10-617-320-3700	Sequence 3700, Ap
29	49	46.7	1497	5	US-10-484-218-18	Sequence 18, Appl
30	47.5	45.2	144	4	US-10-767-701-57788	Sequence 57788, A
31	47	44.8	354	4	US-10-282-122A-46798	Sequence 46798, A
32	47	44.8	787	5	US-10-484-218-16	Sequence 16, Appl
33	46.5	44.3	479	4	US-10-437-963-189793	Sequence 189793, A
34	46	43.8	227	4	US-10-424-599-273936	Sequence 273936, A
35	46	43.8	342	4	US-10-369-493-19927	Sequence 19927, A
36	46	43.8	402	4	US-10-724-972A-5214	Sequence 5214, Ap
37	45	42.9	399	5	US-10-732-923-10174	Sequence 10174, A
38	45	42.9	584	3	US-09-995-749A-12	Sequence 12, Appl
39	45	42.9	816	4	US-10-282-122A-52967	Sequence 52967, A
40	45	42.9	890	4	US-10-437-963-171378	Sequence 171378, A
41	45	42.9	2057	4	US-10-417-280A-2	Sequence 2, Appl
42	44.5	42.4	415	5	US-10-739-930-8402	Sequence 8402, Ap
43	44.5	42.4	449	4	US-10-425-114-46598	Sequence 46598, A
44	44.5	42.4	449	4	US-10-425-114-61364	Sequence 61364, A
45	44.5	42.4	468	4	US-10-425-115-212787	Sequence 212787, A

ALIGNMENTS

RESULT 1

US-10-797-821-43
; Sequence 43, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 43
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus

US-10-797-821-43

Query Match 100.0%; Score 105; DB 5; Length 20;
Best Local Similarity 100.0%; Pred. No. 3.2e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVIANNVDFVSWGITDFEM 20
|||
DB 1 VVIANNVDFVSWGITDFEM 20

RESULT 2

US-10-383-930-37
; Sequence 37, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930

```

; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 37
; LENGTH: 1590
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-383-930-37

Query Match      100.0%; Score 105; DB 4; Length 1590;
Best Local Similarity 100.0%; Pred. No. 2.9e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 VVIANNVDFVSWGITDFEM 20
      |||||
Db      847 VVIANNVDFVSWGITDFEM 866

RESULT 3
US-10-797-821-37
; Sequence 37, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 37
; LENGTH: 1590
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-797-821-37

Query Match      100.0%; Score 105; DB 5; Length 1590;
Best Local Similarity 100.0%; Pred. No. 2.9e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 VVIANNVDFVSWGITDFEM 20
      |||||
Db      847 VVIANNVDFVSWGITDFEM 866

RESULT 4
US-09-740-274-4
; Sequence 4, Application US/09740274
; Patent No. US20020031826A1
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361

```

```

; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-4

Query Match      85.7%; Score 90; DB 3; Length 1375;
Best Local Similarity 80.0%; Pred. No. 5.4e-05;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      1 VVIANNVDFVSWGITDFEM 20
      |||||
Db      876 VVIANNVDFVSWGITDFEM 895

RESULT 5
US-10-383-930-35
; Sequence 35, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 35
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-35

Query Match      85.7%; Score 90; DB 4; Length 1375;
Best Local Similarity 80.0%; Pred. No. 5.4e-05;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      1 VVIANNVDFVSWGITDFEM 20
      |||||
Db      876 VVIANNVDFVSWGITDFEM 895

RESULT 6
US-10-797-821-35
; Sequence 35, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09

```

; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 35
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-35

Query Match 85.7%; Score 90; DB 5; Length 1375;
Best Local Similarity 80.0%; Pred. No. 5.4e-05;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVIANNVDKFSWGVTDFEM 20
||| ||||| ||:||||
Db 876 VVIANNVDKFAEWGVTDFEM 895

RESULT 7

US-09-740-274-2
; Sequence 2, Application US/09740274
; Patent No. US20020031826A1
; GENERAL INFORMATION:

; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-740-274-2

Query Match 85.7%; Score 90; DB 3; Length 1475;
Best Local Similarity 80.0%; Pred. No. 5.8e-05;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVIANNVDKFSWGVTDFEM 20
||| ||||| ||:||||
Db 849 VVIANNVDKFAEWGVTDFEM 868

RESULT 8

US-10-383-930-34
; Sequence 34, Application US/10383930

; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 34
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-34

Query Match 85.7%; Score 90; DB 4; Length 1475;
Best Local Similarity 80.0%; Pred. No. 5.8e-05;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVIANNVDKFSWGVTDFEM 20
||| ||||| ||:||||
Db 849 VVIANNVDKFAEWGVTDFEM 868

RESULT 9

US-10-797-821-34
; Sequence 34, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:

; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 34
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-34

Query Match 85.7%; Score 90; DB 5; Length 1475;
Best Local Similarity 80.0%; Pred. No. 5.8e-05;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVIANNVDKFSWGVTDFEM 20
||| ||||| ||:||||
Db 849 VVIANNVDKFAEWGVTDFEM 868

RESULT 10

US-10-383-930-40
; Sequence 40, Application US/10383930
; Publication No. US20040127400A1

```
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 40
; LENGTH: 1518
; TYPE: PRT
; ORGANISM: Streptococcus salivarius
US-10-383-930-40

Query Match          67.6%; Score 71; DB 4; Length 1518;
Best Local Similarity 72.2%; Pred. No. 0.054;
Matches 13; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 IANNVDKFKVSWGITDFEM 20
||| ||| ||| ||| |||
Db 922 IAEVDLFLKSWGVTSEFEM 939

RESULT 11
US-10-797-821-40
; Sequence 40, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 40
; LENGTH: 1518
; TYPE: PRT
; ORGANISM: Streptococcus salivarius
US-10-797-821-40

Query Match          67.6%; Score 71; DB 5; Length 1518;
Best Local Similarity 72.2%; Pred. No. 0.054;
Matches 13; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 IANNVDKFKVSWGITDFEM 20
||| ||| ||| ||| |||
Db 922 IAEVDLFLKSWGVTSEFEM 939

RESULT 12
US-10-383-930-39
; Sequence 39, Application US/10383930
; Publication No. US2004012400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 39
; LENGTH: 1365
; TYPE: PRT
; ORGANISM: Streptococcus downei
US-10-383-930-39

Query Match          65.7%; Score 69; DB 4; Length 1365;
Best Local Similarity 73.7%; Pred. No. 0.099;
Matches 14; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 VIANNVDKFKVSWGITDFEM 20
||| ||| ||| ||| |||
Db 837 VIAKNGDLFKSWGITQFEM 855

RESULT 13
US-10-797-821-39
; Sequence 39, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 39
; LENGTH: 1365
; TYPE: PRT
; ORGANISM: Streptococcus downei
US-10-797-821-39

Query Match          65.7%; Score 69; DB 5; Length 1365;
Best Local Similarity 73.7%; Pred. No. 0.099;
Matches 14; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 VIANNVDKFKVSWGITDFEM 20
||| ||| ||| ||| |||
Db 837 VIAKNGDLFKSWGITQFEM 855

RESULT 14
US-09-995-749A-11
; Sequence 11, Application US/09995749A
; Patent No. US2002015568A1
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
```

```

; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: B043388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 522
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-09-995-749A-11

```

```

Query Match          64.8%; Score 68; DB 3; Length 522;
Best Local Similarity 55.0%; Pred. No. 0.052;
Matches 11; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

```

```

QY      1  VTIANNVDKFSVSGITDFEM 20
Db      438  VTIANNVDKFSVSGITDFEM 457

```

```

RESULT 15
US-09-995-749A-10
; Sequence 10, Application US/09995749A
; Patent No. US20020155568A1
; GENERAL INFORMATION:
; APPLICANT: VAN GEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: B043388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-995-749A-10

```

```

Query Match          61.9%; Score 65; DB 3; Length 545;
Best Local Similarity 66.7%; Pred. No. 0.16;
Matches 12; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

```

```

QY      3  IANNVDKFSVSGITDFEM 20
Db      463  IANNVDKFSVSGITDFEM 480

```

Search completed: February 11, 2006, 20:57:30
Job time : 2.79012 secs

THIS PAGE BLANK (USPTO)

Result No.	Score			Query Match	Length	DB	ID	Description
	45	5	43					
1	45	5	43	3	574	6	US-10-517-939-108	Sequence 108, App
2	41	39	37	10	604	6	US-10-517-939-34	Sequence 34, Appl
3	40	38	1	1	336	6	US-10-517-939-134	Sequence 134, App
4	39	37	1	180	6	US-10-519-390-20	Sequence 20, Appl	
5	39	37	1	180	7	US-11-176-830-213	Sequence 213, App	
6	39	37	1	180	7	US-11-176-830-729	Sequence 729, App	
7	39	37	1	180	7	US-11-176-830-730	Sequence 730, App	
8	39	37	1	180	7	US-11-176-830-731	Sequence 731, App	
9	39	37	1	180	7	US-11-176-830-732	Sequence 732, App	
10	39	37	1	180	7	US-11-176-830-733	Sequence 733, App	
11	39	37	1	180	7	US-11-176-830-734	Sequence 734, App	
12	39	37	1	180	7	US-11-176-830-735	Sequence 735, App	
13	39	37	1	180	7	US-11-176-830-736	Sequence 736, App	
14	39	37	1	180	7	US-11-176-830-737	Sequence 737, App	
15	39	37	1	180	7	US-11-176-830-738	Sequence 738, App	
16	39	37	1	180	7	US-11-176-830-739	Sequence 739, App	
17	39	37	1	180	7	US-11-176-830-740	Sequence 740, App	
18	39	37	1	180	7	US-11-176-830-741	Sequence 741, App	
19	39	37	1	180	7	US-11-176-830-742	Sequence 742, App	
20	39	37	1	180	7	US-11-176-830-743	Sequence 743, App	
21	39	37	1	180	7	US-11-176-830-744	Sequence 744, App	
22	39	37	1	180	7	US-11-176-830-745	Sequence 745, App	
23	39	37	1	180	7	US-11-176-830-746	Sequence 746, App	
24	39	37	1	180	7	US-11-176-830-747	Sequence 747, App	
25	39	37	1	180	7	US-11-176-830-748	Sequence 748, App	


```
; APPLICANT: Blum, David
; APPLICANT: Esteghlalian, Alireza
; TITLE OF INVENTION: XYLANASES, NUCLEIC ACIDS ENCODING THEM
; FILE REFERENCE: 564462007901
; CURRENT APPLICATION NUMBER: US/10/517,939
; PRIOR FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: PCT/US03/19153
; PRIOR FILING DATE: 2003-06-16
; PRIOR APPLICATION NUMBER: 60/389,299
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 380
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 604
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
; NAME/KEY: SIGNAL
; LOCATION: (1)...(23)
US-10-517-939-34
```

```
Query Match 39.0%; Score 41; DB 6; Length 604;
Best Local Similarity 62.5%; Pred. No. 39;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 10 FVSWGITD 17
|:|:|:|
Db 299 FITWGVTD 306
```

```
RESULT 3
US-10-517-939-134
; Sequence 134, Application US/10517939
; Publication No. US20060003433A1
; GENERAL INFORMATION:
; APPLICANT: Steer, Brian
; APPLICANT: Callen, Walter
; APPLICANT: Healey, Shaun
; APPLICANT: Hazlewood, Geoff
; APPLICANT: Wu, Di
; APPLICANT: Blum, David
; APPLICANT: Esteghlalian, Alireza
; TITLE OF INVENTION: XYLANASES, NUCLEIC ACIDS ENCODING THEM
; FILE REFERENCE: 564462007901
; CURRENT APPLICATION NUMBER: US/10/517,939
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: PCT/US03/19153
; PRIOR FILING DATE: 2003-06-16
; PRIOR APPLICATION NUMBER: 60/389,299
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 380
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 134
; LENGTH: 336
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-517-939-134
```

```
Query Match 38.1%; Score 40; DB 6; Length 336;
Best Local Similarity 53.3%; Pred. No. 31;
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
```

```
QY 5 NNVDKFSWGITDF 19
|:|:|:|
Db 322 NHQPKESFWGIVDF 336
```

```
RESULT 4
US-10-519-390-20
; Sequence 20, Application US/10519390
; Publication No. US2006000872A1
; GENERAL INFORMATION:
; APPLICANT: MEDEXGEN Inc.
; APPLICANT: CHUNG, Yong-Hoon
; APPLICANT: LEE, Hak-sup
; APPLICANT: YI, Ki-Wan
; APPLICANT: KIM, Jae-Youn
; APPLICANT: HEO, Youn-Hwa
; TITLE OF INVENTION: A method of improving efficacy of biological response-modifying
; TITLE OF INVENTION: proteins and the example mutants
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/519,390
; CURRENT FILING DATE: 2004-12-23
; PRIOR APPLICATION NUMBER: KR10-2003-0051846
; PRIOR FILING DATE: 2003-07-26
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: KopatentIn 1.71
; SEQ ID NO 20
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: LIP: 41st, 52nd, 67th, 70th, 156th or 180th Phe is replaced by
; OTHER INFORMATION: Val.
US-10-519-390-20
```

```
Query Match 37.1%; Score 39; DB 6; Length 180;
Best Local Similarity 50.0%; Pred. No. 23;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
```

```
QY 5 NNVDKFSWGITDF 18
|:|:|:|
Db 54 NNLDKLCGPNVTDF 67
```

```
RESULT 5
US-11-176-830-213
; Sequence 213, Application US/11176830
; Publication No. US20060020116A1
; GENERAL INFORMATION:
; APPLICANT: Gantier, Rene
; APPLICANT: Guyon, Thierry
; APPLICANT: Dritanti, Lila
; APPLICANT: Vega, Manuel
; TITLE OF INVENTION: Rational Evolution of Cytokines for Higher Stability, Encoding N
; TITLE OF INVENTION: Acid Molecules and Related Applications
; FILE REFERENCE: 17109-012002 (922B)
; CURRENT APPLICATION NUMBER: US/11/176,830
; CURRENT FILING DATE: 2005-07-06
; PRIOR APPLICATION NUMBER: 10/658,834
; PRIOR FILING DATE: 2003-09-08
; PRIOR APPLICATION NUMBER: 60/457,135
; PRIOR FILING DATE: 2003-03-21
; PRIOR APPLICATION NUMBER: 60/409,898
; PRIOR FILING DATE: 2002-09-09
; NUMBER OF SEQ ID NOS: 1306
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 213
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: Genbank CAA32147
; DATABASE ENTRY DATE: 1995-03-22
US-11-176-830-213
```

```
Query Match 37.1%; Score 39; DB 7; Length 180;
Best Local Similarity 50.0%; Pred. No. 23;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
```

QY 5 NNVDKFSWGITDF 18
 ||:||:||||
 Db 54 NNLDKLCGPNVTDF 67

RESULT 6

US-11-176-830-729
 ; Sequence 729, Application US/11176830
 ; Publication No. US20060020116A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Gantier, Rene
 ; APPLICANT: Guyon, Thierry
 ; APPLICANT: Drittanti, Lila
 ; APPLICANT: Vega, Manuel
 ; TITLE OF INVENTION: Rational Evolution of Cytokines for Higher Stability, Encoding Nu
 ; TITLE OF INVENTION: Acid Molecules and Related Applications
 ; FILE REFERENCE: 17109-012002 (922B)
 ; CURRENT APPLICATION NUMBER: US/11/176,830
 ; CURRENT FILING DATE: 2005-07-06
 ; PRIOR APPLICATION NUMBER: 10/658,834
 ; PRIOR FILING DATE: 2003-09-08
 ; PRIOR APPLICATION NUMBER: 60/457,135
 ; PRIOR FILING DATE: 2003-03-21
 ; PRIOR APPLICATION NUMBER: 60/409,898
 ; PRIOR FILING DATE: 2002-09-09
 ; NUMBER OF SEQ ID NOS: 1306
 ; SOFTWARE: FastSEQ for Windows Version 4.0
 ; SEQ ID NO 729
 ; LENGTH: 180
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-11-176-830-729

Query Match 37.1%; Score 39; DB 7; Length 180;
 Best Local Similarity 50.0%; Pred. No. 23;
 Matches 7; Conservative 2; Mismatches 5; Indels 5; Gaps 0;

QY 5 NNVDKFSWGITDF 18
 ||:||:||||
 Db 54 NNLDKLCGPNVTDF 67

RESULT 7

US-11-176-830-730
 ; Sequence 730, Application US/11176830
 ; Publication No. US20060020116A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Gantier, Rene
 ; APPLICANT: Guyon, Thierry
 ; APPLICANT: Drittanti, Lila
 ; APPLICANT: Vega, Manuel
 ; TITLE OF INVENTION: Rational Evolution of Cytokines for Higher Stability, Encoding Nu
 ; TITLE OF INVENTION: Acid Molecules and Related Applications
 ; FILE REFERENCE: 17109-012002 (922B)
 ; CURRENT APPLICATION NUMBER: US/11/176,830
 ; CURRENT FILING DATE: 2005-07-06
 ; PRIOR APPLICATION NUMBER: 10/658,834
 ; PRIOR FILING DATE: 2003-09-08
 ; PRIOR APPLICATION NUMBER: 60/457,135
 ; PRIOR FILING DATE: 2003-03-21
 ; PRIOR APPLICATION NUMBER: 60/409,898
 ; PRIOR FILING DATE: 2002-09-09
 ; NUMBER OF SEQ ID NOS: 1306
 ; SOFTWARE: FastSEQ for Windows Version 4.0
 ; SEQ ID NO 730
 ; LENGTH: 180
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-11-176-830-730

Query Match 37.1%; Score 39; DB 7; Length 180;
 Best Local Similarity 50.0%; Pred. No. 23;
 Matches 7; Conservative 2; Mismatches 5; Indels 5; Gaps 0;

QY 5 NNVDKFSWGITDF 18
 ||:||:||||
 Db 54 NNLDKLCGPNVTDF 67

RESULT 8

US-11-176-830-731
 ; Sequence 731, Application US/11176830
 ; Publication No. US20060020116A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Gantier, Rene
 ; APPLICANT: Guyon, Thierry
 ; APPLICANT: Drittanti, Lila
 ; APPLICANT: Vega, Manuel
 ; TITLE OF INVENTION: Rational Evolution of Cytokines for Higher Stability, Encoding N
 ; TITLE OF INVENTION: Acid Molecules and Related Applications
 ; FILE REFERENCE: 17109-012002 (922B)
 ; CURRENT APPLICATION NUMBER: US/11/176,830
 ; CURRENT FILING DATE: 2005-07-06
 ; PRIOR APPLICATION NUMBER: 10/658,834
 ; PRIOR FILING DATE: 2003-09-08
 ; PRIOR APPLICATION NUMBER: 60/457,135
 ; PRIOR FILING DATE: 2003-03-21
 ; PRIOR APPLICATION NUMBER: 60/409,898
 ; PRIOR FILING DATE: 2002-09-09
 ; NUMBER OF SEQ ID NOS: 1306
 ; SOFTWARE: FastSEQ for Windows Version 4.0
 ; SEQ ID NO 731
 ; LENGTH: 180
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-11-176-830-731

Query Match 37.1%; Score 39; DB 7; Length 180;
 Best Local Similarity 50.0%; Pred. No. 23;
 Matches 7; Conservative 2; Mismatches 5; Indels 5; Gaps 0;

QY 5 NNVDKFSWGITDF 18
 ||:||:||||
 Db 54 NNLDKLCGPNVTDF 67

RESULT 9

US-11-176-830-732
 ; Sequence 732, Application US/11176830
 ; Publication No. US20060020116A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Gantier, Rene
 ; APPLICANT: Guyon, Thierry
 ; APPLICANT: Drittanti, Lila
 ; APPLICANT: Vega, Manuel
 ; TITLE OF INVENTION: Rational Evolution of Cytokines for Higher Stability, Encoding N
 ; TITLE OF INVENTION: Acid Molecules and Related Applications
 ; FILE REFERENCE: 17109-012002 (922B)
 ; CURRENT APPLICATION NUMBER: US/11/176,830
 ; CURRENT FILING DATE: 2005-07-06
 ; PRIOR APPLICATION NUMBER: 10/658,834
 ; PRIOR FILING DATE: 2003-09-08
 ; PRIOR APPLICATION NUMBER: 60/457,135
 ; PRIOR FILING DATE: 2003-03-21
 ; PRIOR APPLICATION NUMBER: 60/409,898
 ; PRIOR FILING DATE: 2002-09-09
 ; NUMBER OF SEQ ID NOS: 1306
 ; SOFTWARE: FastSEQ for Windows Version 4.0
 ; SEQ ID NO 732
 ; LENGTH: 180
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-11-176-830-732

Query Match 37.1%; Score 39; DB 7; Length 180;
 Best Local Similarity 50.0%; Pred. No. 23;

```
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 5 NNVDKFSVSGITDF 18
Db 54 NNLDKLCGPNVTDF 67

RESULT 10
US-11-176-830-733
; Sequence 733, Application US/11176830
; Publication No. US20060020116A1
; GENERAL INFORMATION:
; APPLICANT: Gantier, Rene
; APPLICANT: Guyon, Thierry
; APPLICANT: Drittanti, Lila
; APPLICANT: Vega, Manuel
; TITLE OF INVENTION: Rational Evolution of Cytokines for Higher Stability, Encoding N
; FILE REFERENCE: 17109-012002 (922B)
; CURRENT APPLICATION NUMBER: US/11/176,830
; CURRENT FILING DATE: 2005-07-06
; PRIOR FILING DATE: 2003-09-08
; PRIOR APPLICATION NUMBER: 10/658,834
; PRIOR FILING DATE: 2003-09-08
; PRIOR APPLICATION NUMBER: 60/457,135
; PRIOR FILING DATE: 2003-03-21
; PRIOR APPLICATION NUMBER: 60/409,898
; PRIOR FILING DATE: 2002-09-09
; NUMBER OF SEQ ID NOS: 1306
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 733
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-176-830-733

Query Match 37.1%; Score 39; DB 7; Length 180;
Best Local Similarity 50.0%; Pred. No. 23;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 5 NNVDKFSVSGITDF 18
Db 54 NNLDKLCGPNVTDF 67

RESULT 11
US-11-176-830-734
; Sequence 734, Application US/11176830
; Publication No. US20060020116A1
; GENERAL INFORMATION:
; APPLICANT: Gantier, Rene
; APPLICANT: Guyon, Thierry
; APPLICANT: Drittanti, Lila
; APPLICANT: Vega, Manuel
; TITLE OF INVENTION: Rational Evolution of Cytokines for Higher Stability, Encoding N
; FILE REFERENCE: 17109-012002 (922B)
; CURRENT APPLICATION NUMBER: US/11/176,830
; CURRENT FILING DATE: 2005-07-06
; PRIOR FILING DATE: 2003-09-08
; PRIOR APPLICATION NUMBER: 10/658,834
; PRIOR FILING DATE: 2003-09-08
; PRIOR APPLICATION NUMBER: 60/457,135
; PRIOR FILING DATE: 2003-03-21
; PRIOR APPLICATION NUMBER: 60/409,898
; PRIOR FILING DATE: 2002-09-09
; NUMBER OF SEQ ID NOS: 1306
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 734
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-176-830-734

Query Match 37.1%; Score 39; DB 7; Length 180;
```

```
Best Local Similarity 50.0%; Pred. No. 23;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 5 NNVDKFSVSGITDF 18
Db 54 NNLDKLCGPNVTDF 67

RESULT 12
US-11-176-830-735
; Sequence 735, Application US/11176830
; Publication No. US20060020116A1
; GENERAL INFORMATION:
; APPLICANT: Gantier, Rene
; APPLICANT: Guyon, Thierry
; APPLICANT: Drittanti, Lila
; APPLICANT: Vega, Manuel
; TITLE OF INVENTION: Rational Evolution of Cytokines for Higher Stability, Encoding N
; FILE REFERENCE: 17109-012002 (922B)
; CURRENT APPLICATION NUMBER: US/11/176,830
; CURRENT FILING DATE: 2005-07-06
; PRIOR FILING DATE: 2003-09-08
; PRIOR APPLICATION NUMBER: 10/658,834
; PRIOR FILING DATE: 2003-09-08
; PRIOR APPLICATION NUMBER: 60/457,135
; PRIOR FILING DATE: 2003-03-21
; PRIOR APPLICATION NUMBER: 60/409,898
; PRIOR FILING DATE: 2002-09-09
; NUMBER OF SEQ ID NOS: 1306
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 735
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-176-830-735

Query Match 37.1%; Score 39; DB 7; Length 180;
Best Local Similarity 50.0%; Pred. No. 23;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 5 NNVDKFSVSGITDF 18
Db 54 NNLDKLCGPNVTDF 67

RESULT 13
US-11-176-830-736
; Sequence 736, Application US/11176830
; Publication No. US20060020116A1
; GENERAL INFORMATION:
; APPLICANT: Gantier, Rene
; APPLICANT: Guyon, Thierry
; APPLICANT: Drittanti, Lila
; APPLICANT: Vega, Manuel
; TITLE OF INVENTION: Rational Evolution of Cytokines for Higher Stability, Encoding N
; FILE REFERENCE: 17109-012002 (922B)
; CURRENT APPLICATION NUMBER: US/11/176,830
; CURRENT FILING DATE: 2005-07-06
; PRIOR FILING DATE: 2003-09-08
; PRIOR APPLICATION NUMBER: 10/658,834
; PRIOR FILING DATE: 2003-03-21
; PRIOR APPLICATION NUMBER: 60/409,898
; PRIOR FILING DATE: 2002-09-09
; NUMBER OF SEQ ID NOS: 1306
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 736
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-176-830-736
```

Query Match 37.1%; Score 39; DB 7; Length 180;
Best Local Similarity 50.0%; Pred. No. 23;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 5 NNVDKFSWGITDF 18
Db 54 NNLDKLCGPNVTDF 67

RESULT 14

US-11-176-830-737
; Sequence 737, Application US/11176830
; Publication No. US20060020116A1
; GENERAL INFORMATION:
; APPLICANT: Gantier, Rene
; APPLICANT: Guyon, Thierry
; APPLICANT: Drihtanti, Lila
; APPLICANT: Vega, Manuel
; TITLE OF INVENTION: Rational Evolution of Cytokines for Higher Stability, Encoding Nu
; TITLE OF INVENTION: Acid Molecules and Related Applications
; FILE REFERENCE: 17109-012002 (922B)
; CURRENT APPLICATION NUMBER: US/11/176,830
; CURRENT FILING DATE: 2005-07-06
; PRIOR APPLICATION NUMBER: 10/658,834
; PRIOR FILING DATE: 2003-09-08
; PRIOR APPLICATION NUMBER: 60/457,135
; PRIOR FILING DATE: 2003-03-21
; PRIOR APPLICATION NUMBER: 60/409,898
; PRIOR FILING DATE: 2002-09-09
; NUMBER OF SEQ ID NOS: 1306
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 737
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-176-830-737

Query Match 37.1%; Score 39; DB 7; Length 180;
Best Local Similarity 50.0%; Pred. No. 23;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 5 NNVDKFSWGITDF 18
Db 54 NNLDKLCGPNVTDF 67

RESULT 15

US-11-176-830-738
; Sequence 738, Application US/11176830
; Publication No. US20060020116A1
; GENERAL INFORMATION:
; APPLICANT: Gantier, Rene
; APPLICANT: Guyon, Thierry
; APPLICANT: Drihtanti, Lila
; APPLICANT: Vega, Manuel
; TITLE OF INVENTION: Rational Evolution of Cytokines for Higher Stability, Encoding Nu
; TITLE OF INVENTION: Acid Molecules and Related Applications
; FILE REFERENCE: 17109-012002 (922B)
; CURRENT APPLICATION NUMBER: US/11/176,830
; CURRENT FILING DATE: 2005-07-06
; PRIOR APPLICATION NUMBER: 10/658,834
; PRIOR FILING DATE: 2003-09-08
; PRIOR APPLICATION NUMBER: 60/457,135
; PRIOR FILING DATE: 2003-03-21
; PRIOR APPLICATION NUMBER: 60/409,898
; PRIOR FILING DATE: 2002-09-09
; NUMBER OF SEQ ID NOS: 1306
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 738
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-176-830-738

Query Match 37.1%; Score 39; DB 7; Length 180;
Best Local Similarity 50.0%; Pred. No. 23;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 5 NNVDKFSWGITDF 18
Db 54 NNLDKLCGPNVTDF 67

Search completed: February 11, 2006, 20:59:03
Job time : 1.11381 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 19:15:50 ; Search time 0.412809 Seconds
(without alignments)
4661.567 Million cell updates/sec

Title: US-10-797-821-43

Perfect score: 105
Sequence: 1 VVIANNVDKFSWGITDFEM 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 80.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	105	100.0	1592	2 A38175	glucosyltransferas
2	90	85.7	1375	2 JTO345	dextranucrase (EC
3	90	85.7	1475	2 B33135	gtfB protein precu
4	71	67.6	1518	2 A44811	glucosyltransferas
5	69	65.7	1365	2 A41483	glucosyltransferas
6	65	61.9	1431	2 A45866	dextranucrase (EC
7	64	61.0	1290	2 JCS473	dextranucrase (EC
8	63	60.0	1599	2 S22737	glucosyltransferas
9	58	55.2	1449	2 T30857	glucosyltransferas
10	58	55.2	1449	2 T30552	glucosyltransferas
11	56	53.3	1577	2 T30858	glucosyltransferas
12	55	52.4	382	2 AH2361	endo-1,4-beta-xyla
13	53	50.5	1508	2 T31098	probable dextranu
14	51	48.6	459	2 G97906	transcription anti
15	49	46.7	493	2 B95036	transcription regu
16	47	44.8	178	2 S44624	C50C3.5 protein -
17	47	44.8	396	2 T11709	nifs protein homol
18	46	43.8	300	2 I40425	carboxylesterase (
19	45.5	43.3	837	1 A31842	endo-1,4-beta-xyla
20	45	42.9	200	2 T47531	hypothetical prote
21	44	41.9	412	2 T42024	probable serine pr
22	44	41.9	810	2 JC4837	hypoxia-inducible
23	43.5	41.4	133	2 AE2629	apaGB protein (impo
24	43.5	41.4	133	2 C97411	apaGB protein (AJO
25	43	41.0	116	2 F82805	conserved hypothet
26	43	41.0	167	2 H81252	NADH2 dehydrogenas
27	43	41.0	271	2 T37745	CAAX prenyl protei
28	43	41.0	295	2 I64050	probable integral
29	43	41.0	304	2 A49496	malate dehydrogena

30	43	41.0	456	2 AH2778	ABC transporter, s
31	43	41.0	456	2 D97558	hypothetical prote
32	43	41.0	813	2 JCS809	hypoxia-inducible
33	43	41.0	826	2 I38972	hypoxia-inducible
34	42.5	40.5	301	2 C84474	hypothetical prote
35	42.5	40.5	1136	2 T40355	hypothetical prote
36	42	40.0	85	2 H81049	hypothetical prote
37	42	40.0	146	2 C90454	hypothetical prote
38	42	40.0	165	2 G88473	protein F40H6.1 (i
39	42	40.0	200	2 E81314	probable periplasm
40	42	40.0	242	2 F97158	hypothetical prote
41	42	40.0	331	2 AE3372	toluene tolerance
42	42	40.0	334	2 AD1049	HfIC protein (EC 3
43	42	40.0	334	2 G86113	probable integral
44	42	40.0	334	2 G91272	probable integral
45	42	40.0	334	2 C43653	probable integral

ALIGNMENTS

RESULT 1

A38175

Glucosyltransferase precursor - Streptococcus sobrinus

C:Species: Streptococcus sobrinus

C:Date: 28-Aug-1992 #sequence_revision 28-Aug-1992 #text_change 15-Mar-2004

C:Accession: A38175

R;Abo, H.; Mtsamura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.

J. Bacteriol. 173, 989-996, 1991

A:Title: Peptide sequences for sucrose splitting and glucan binding within Streptococcus

A:Reference number: A38175; MUID:91123227; PMID:1704006

A:Accession: A38175

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1592 <ABO>

A:Cross-references: UNIPARC:UPI00012BCB2; GB:D90213; NID:G217032; PIDN:BAA14241.1; PID

F:1093-1112/Domain: cpl repeat homology <CP1>

F:1222-1241/Domain: cpl repeat homology <CP2>

F:1287-1306/Domain: cpl repeat homology <CP3>

F:1330-1351/Domain: cpl repeat homology <CP4>

F:1352-1371/Domain: cpl repeat homology <CP5>

F:1402-1420/Domain: cpl repeat homology <CP6>

F:1465-1484/Domain: cpl repeat homology <CP7>

F:1513-1532/Domain: cpl repeat homology <CP8>

Query Match 100.0%; Score 105; DB 2; Length 1592;

Best Local Similarity 100.0%; Pred. No. 5.8e-08;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVIANNVDKFSWGITDFEM 20

|||||

Db 847 VVIANNVDKFSWGITDFEM 866

RESULT 2

JTO345

dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)

N:Alternate names: sucrose 6-glucosyltransferase

C:Species: Streptococcus mutans

C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 09-Jul-2004

C:Accession: JTO345; C33135

R;Ueda, S.; Shiroza, T.; Kuramitsu, H.K.

Gene 69, 101-109, 1988

A:Title: Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.

A:Reference number: JTO345; MUID:89137980; PMID:2976010

A:Accession: JTO345

A:Molecule type: DNA

A:Residues: 1-1375 <UED>

A:Cross-references: UNIPROT:P13470; UNIPARC:UPI0000155515

A:Experimental source: GS-5

R;Shiroza, T.; Ueda, S.; Kuramitsu, H.K.

J. Bacteriol. 169, 4263-4270, 1987

A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.

A:Reference number: A33135; MUID:87308013; PMID:3040685
A:Accession: C33135
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-349 <SHI>
A:Cross-references: UNIPARC:UPI000014E25D; GB:M17361
C:Genetics:
A:Gene: gtfc
C:Function:

A:Description: catalyzes the synthesis of both water-soluble and water-insoluble glucans
C:Keywords: duplication; glycosyltransferase; hexosyltransferase
F:1-34/Domain: signal sequence #status predicted <SIG>
F:35-1375/Product: glucosyltransferase #status predicted <MAT>
F:1126-1145/Domain: cpl repeat homology <CP1>
F:1253-1272/Domain: cpl repeat homology <CP2>
F:1318-1337/Domain: cpl repeat homology <CP3>

Query Match 85.7%; Score 90; DB 2; Length 1375;
Best Local Similarity 80.0%; Pred. No. 1.1e-05;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVIANNVDKFSVSGITDFEM 20
||||| ||||| ||:|||||
Db 876 VVIANNVDKFAENGVTDFEM 895

RESULT 3

B33135
gtfB protein precursor - Streptococcus mutans
C:Species: Streptococcus mutans
C:Date: 23-Oct-1990 #sequence_revision 23-Oct-1990 #text_change 09-Jul-2004
C:Accession: B33135; A33128
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
J. Bacteriol. 169, 4263-4270, 1987
A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.
A:Reference number: A33135; MUID:87308013; PMID:3040685
A:Accession: B33135
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1475 <SHI>
A:Cross-references: UNIPROT:P08987; UNIPARC:UPI000014D972; GB:M17361; NID:g153639; PIDN:
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
submitted to the Protein Sequence Database, September 1990
A:Reference number: A33128
A:Accession: A33128
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-171,173-641,'N',643-1475 <SH2>
A:Cross-references: UNIPARC:UPI000017AC5E
A:Experimental source: strain GS-5
F:1096-1115/Domain: cpl repeat homology <CP1>
F:1224-1243/Domain: cpl repeat homology <CP2>
F:1289-1308/Domain: cpl repeat homology <CP3>
F:1354-1373/Domain: cpl repeat homology <CP4>
F:1419-1438/Domain: cpl repeat homology <CP5>

Query Match 85.7%; Score 90; DB 2; Length 1475;
Best Local Similarity 80.0%; Pred. No. 1.2e-05;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVIANNVDKFSVSGITDFEM 20
||||| ||||| ||:|||||
Db 849 VVIANNVDKFAENGVTDFEM 868

RESULT 4

A4811
glucosyltransferase (EC 2.4.1.-) I - Streptococcus salivarius
C:Species: Streptococcus salivarius
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C:Accession: A44811; S22726; S28809
R:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.
J. Gen. Microbiol. 137, 2577-2593, 1991

A:Title: Molecular characterization of a cluster of at least two glucosyltransferase genes
A:Reference number: A44811; MUID:92148377; PMID:1838391
A:Accession: A44811
A:Molecule type: DNA
A:Residues: 1-1518 <GIF>
A:Cross-references: UNIPROT:Q00600; UNIPARC:UPI00000BEF31; EMBL:Z11873; NID:g47526; PIDN:
A>Note: sequence extracted from NCBI backbone (NCBIN:81050, NCBIP:81052)
C:Genetics:
A:Gene: gtff
C:Function:

A:Description: glycosyltransferase; hexosyltransferase
F:1307-1326/Domain: cpl repeat homology <CP4>

Query Match 67.6%; Score 71; DB 2; Length 1518;
Best Local Similarity 72.2%; Pred. No. 0.011;
Matches 13; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 IANNVDKFSVSGITDFEM 20
||||| ||||| ||:|||||
Db 922 IANNVDLFSKSGVTSEFEM 939

RESULT 5

A41483
glucosyltransferase (EC 2.4.1.-) gtfS precursor - Streptococcus sobrinus
C:Species: Streptococcus sobrinus
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 15-Mar-2004
C:Accession: A41483
R:Gilmore, K.S.; Russell, R.R.B.; Ferretti, J.J.
Infect. Immun. 58, 2452-2458, 1990
A:Title: Analysis of the Streptococcus downei gtfS gene, which specifies a glucosyltran:
A:Reference number: A41483; MUID:90316665; PMID:2142479
A:Accession: A41483
A:Molecule type: DNA
A:Residues: 1-1365 <GIL>
A:Cross-references: UNIPARC:UPI000012BCB6; GB:M30943; NID:g153652; PIDN:AAA36898.1; PID
C:Genetics:
A:Gene: gtfS
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 65.7%; Score 69; DB 2; Length 1365;
Best Local Similarity 73.7%; Pred. No. 0.02;
Matches 14; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 VIANNVDKFSVSGITDFEM 20
||||| ||||| ||:|||||
Db 837 VIAKNGDLFSKSGITQFEM 855

RESULT 6

A45866
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans
C:Species: Streptococcus mutans
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Mar-2004
C:Accession: A45866
R:Honda, O.; Kato, C.; Kuramitsu, H.K.
J. Gen. Microbiol. 136, 2099-2105, 1990
A:Title: Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the glucosyl
A:Reference number: A45866; MUID:91100958; PMID:2148600
A:Accession: A45866
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1431 <HON>
A:Cross-references: UNIPARC:UPI000017AC5C; GB:M29296
C:Keywords: glycosyltransferase; hexosyltransferase
F:181-201/Domain: cpl repeat homology <CP1>
F:1127-1146/Domain: cpl repeat homology <CP2>
F:1192-1211/Domain: cpl repeat homology <CP3>
F:1257-1276/Domain: cpl repeat homology <CP4>
F:1277-1297/Domain: cpl repeat homology <CP5>
F:1321-1340/Domain: cpl repeat homology <CP8>
F:1341-1361/Domain: cpl repeat homology <CP6>
F:1385-1404/Domain: cpl repeat homology <CP7>

```

Query Match          †      61.9%; Score 65; DB 2; Length 1431;
Best Local Similarity 66.7%; Pred.No. 0.087;
Matches 12; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY   3 IANNVDKFSWGIIIDFEM 20
DB   883 IAQNVLFKSWGIVTSFEM 900
      ||| | | | | | | | | |
      ||| | | | | | | | | |

RESULT 7
JCS473
dextranucrase (EC 2.4.1.5) - Leuconostoc mesenteroides
C;Species: Leuconostoc mesenteroides
C;Date: 07-Jul-1997 #sequence_revision 29-Aug-1997 #text_change 09-Jul-2004
C;Accession: JCS473
R;Monchois, V.; Willemot, R.M.; Remaud-Simeon, M.; Croux, C.; Monsan, P.
Gene 182, 23-32, 1996
A;Title: Cloning and sequencing of a gene coding for a novel dextranucrase from Leuconostoc
A;Reference number: JCS473; MUID: 9713688; PMID: 8982063
A;Accession: JCS473
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1290 <MON>
A;Cross-references: UNIPROT: Q48756; UNIPARC: UPI000017ABDF; GB: U38181
C;Comment: This enzyme catalyzes the transfer of D-glucopyranosyl units from sucrose onto
C;Genetics:
A;Gene: dara
C;Keywords: glycosyltransferase; hexosyltransferase
F;78-870/Domain: catalytic #status predicted <CAT>
F;922-1290/Domain: glucan-binding #status predicted <CCB>

Query Match          61.0%; Score 64; DB 2; Length 1290;
Best Local Similarity 60.0%; Pred.No. 0.11;
Matches 12; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY   1 VVIANNVDFKVSWGIIIDFEM 20
DB   686 IKIAQNVNLFKDGIITSFEM 705
      : ||||: | |||| |||
      †

RESULT 8
RS2737
glucosyltransferase (EC 2.4.1.-) S - Streptococcus salivarius
C;Species: Streptococcus salivarius
C;Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 09-Jul-2004
C;Accession: RS2737; S28810; B44811; S22727
R;Jacques, N.
submitted to the EMBL Data Library, March 1992
A;Reference number: S22726
A;Accession: RS2737
A;Molecule type: DNA
A;Residues: 1-1599 <JAC>
A;Cross-references: UNIPROT: Q00599; UNIPARC: UPI00000BEF34; EMBL: Z11872; NID: g47530; PIDN:
A;Experimental source: ATCC 25975
R;Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.
J. Gen. Microbiol. 137, 2577-2593, 1991
A;Title: Molecular characterization of a cluster of at least two glucosyltransferase gen
A;Reference number: A44811; MUID: 92148377; PMID: 1838391
A;Accession: S28810
A;Molecule type: DNA
A;Residues: 1-51 <GIF>
A;Cross-references: UNIPARC: UPI000017027C; EMBL: Z11873
C;Genetics:
A;Gene: gtfk
C;Keywords: glycosyltransferase; hexosyltransferase
F;1456-1475/Domain: cpl repeat homology <CPR>

Query Match          60.0%; Score 63; DB 2; Length 1599;
Best Local Similarity 63.2%; Pred.No. 0.2;
Matches 12; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY   2 VVIANNVDFKVSWGIIIDFEM 20

```


A;Gene: qtfm

Query Match 53.3%; Score 56; DB 2; Length 1577;
Best Local Similarity 61.1%; Pred. No. 2.4;
Matches 11; Conservative 0; Mismatches 7; Indels

Qy 2 VIANNVDKFVSWGITDFE 19
||| | | | | |
Db 958 VIAONAKLFKEWGITSFE 975

RESULT 12

AH2361
endo-1,4-beta-xylanase [imported] - Nostoc sp. (strain PCC 7120)
C|Species: Nostoc sp. PCC 7120
A|Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C|Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 31-Dec-2004
C|Accession: AH2361
R|Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchihara, M.; Tanaka, K.; Shimizu, T.; Ohmura, Y.; Ogiwara, H.; Nakazaki, N.; Shimpō, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.; Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchihara, M.; Tanaka, K.; Shimizu, T.; Ohmura, Y.; Ogiwara, H.; Nakazaki, N.; Shimpō, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
D|Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120
A|Reference number: AB1807; MUID:21595285; PMID:11759840
A|Accession: AH2361
A|Status: preliminary
A|Molecule type: DNA
A|Cross-references: 1-382 <KUR>
A|Experimental source: strain PCC 7120
C|Genetics:
C|Gene: alx4448
C|Superfamily: xylanase; Streptomyces endo-1,4-beta-xylanase A homology

Query Match 52.4%; Score 55; DB 2; Length 382;
Best Local Similarity 64.3%; Pred. No. 0.77;
Matches 9: Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 4 ANNVDFVSWGTD 17
||| :|||
pb 315 ANGVDTIITWGTD 328

RESULT 13

T31098
Probable dextranase (EC 2.4.1.5), extracellular - *Leuconostoc mesenteroides*
C:Species: *Leuconostoc mesenteroides*
C:Date: 22-Oct-1999#sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T31098
R:Monchois, V.; Renaud-Simeon, M.; Monsan, P.; Willemot, R.M.
FEMS Microbiol. Lett. 159, 307-315, 1998
A:Title: Cloning and sequencing of a gene coding for an extracellular dextranase (DS)
A:Reference number: Z20981; MUID:98164374; PMID:9503626
A:Accession: T31098
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1508 <NON>
A:Cross-references: UNIPROT:O52224; UNIPARC:UPI00000BB69B; EMBL:AF030129; NID:g27666611;
A:Experimental source: strain NRRL B-1299
C:Genetics:
A:Gene: dsxB
C:Function:
A:Description: produces dextran composed only of alpha (1-6) glucosidic bonds
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 50.5%; Score 53; DB 2; Length 1508;
Best Local Similarity 52.6%; Pred. No. 6.8;
Matches 10: Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 2 VIANNVDFVSWGITDFEM 20
||| | | ||| |::
pb 927 VIAKNGOLEKDWGITSFOL 94

RESULT 14

G97906
transcription antiterminator BglG family bglG [imported] - Streptococcus pneumoniae (str
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C:Accession: G97906
R.; Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczyk, L.; Burett, S.; DeHoff, B.S.; F
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; F
V, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A37872; MUID:21429245; PMID:11544234
A:Accession: G97906
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-659 <KUR>
A:Cross-references: UNIPROT:Q8DR86; UNIPARC:UPI00000E33FC; GB:AE007317; PIDN:AAK9083.1,
C:Genetics:
C:Gene: bglG

Query Match	48.6%;	Score 51;	DB 2;	Length 659;
Best Local Similarity	50.0%;	Pred. No. 5.8;		
Matches	9:	Conservative	4:	Mismatches
			5:	Indels
			0:	Gaps

QY 3 IANNVDKFVSWGITDFEM 20
| | : | | : | | :
pb 371 TIANNMDKEI FWSISDDET 388

DEPT. T 15

transcription regulator, probable [imported] - Streptococcus pneumoniae (strain TIGR4)
 C:Species: Streptococcus pneumoniae
 C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 09-Jul-2004
 C:Accession: B95036
 R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hein
 on, J.D.; Mayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
 nson, T.; Hickey, E.K.; Holt, I.E.
 Science 293, 498-506, 2001
 A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
 A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
 A:Reference number: A95000; MUID:21357209; PMID:11463916
 A:Accession: B95036
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-493 <KUR>
 A:Cross-references: UNIPROT:Q97SM4; UNIPARC:UPI00000513AC; GB:AE005672; PIDN:AAK74483.1.
 A:Experimental source: strain TIGR4
 C:Genetics:
 A:Gene: SP0306

Query Match	46.7%;	Score 49;	DB 2;	Length 493;
Best Local Similarity	50.0%;	Pred. No. 8.7;		
Matches	9.	Conservative	4.	Mismatches
			5.	Indels
				Gaps 0;

QY 3 IANVDFKFSWGITDFEM 20
| | : | | : | | :
271 TI ANMOKPI EWITISDDEY 198

Search completed: February 11, 2006, 19:43:00
Job time : 1.41281 secs

GenCore version 5.1.7
Cpopyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 18:57:35 ; Search time 2.69869 Seconds
(without alignments)
5228.676 Million cell updates/sec

Title: US-10-797-821-43

Perfect score: 105

Sequence: 1 VVIANNVDKFSWGITDFEM 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Uniprot 05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	105	100.0	1590	2 Q55263_9STRE	Q55263 streptococc
2	105	100.0	1590	2 Q59983_9STRE	Q59983 streptococc
3	105	100.0	1592	1 GTF2_STRDO	P27470 streptococc
4	105	100.0	1597	1 GTF1_STRDO	P11001 streptococc
5	90	85.7	1591	2 Q8VUH3_STRMU	Q8vuh3 streptococc
6	90	85.7	1455	1 GTF3_STRMU	P13470 streptococc
7	90	85.7	1476	1 GTFB_STRMU	P08987 streptococc
8	72	68.6	1338	2 Q9WXJ4_9STRE	Q9wxj4 streptococc
9	72	68.6	2835	2 Q8G9Q2_LEUME	Q8g9q2 leuconostoc
10	71	67.6	1518	2 Q00600_STRSL	Q00600 streptococc
11	69	65.7	1016	2 Q9LCJ7_LEUME	Q9lcj7 leuconostoc
12	69	65.7	1365	1 GTF5_STRDO	P29336 streptococc
13	69	65.7	1454	2 Q6TXV4_LEUME	Q6txv4 leuconostoc
14	68	64.8	1522	2 Q6XKX4_LEUME	Q6xkx4 leuconostoc
15	68	64.8	1527	2 Q8KRE1_LEUME	Q8krel leuconostoc
16	68	64.8	1527	2 Q9ZAR4_LEUME	Q9zar4 leuconostoc
17	65	61.9	1462	1 GTFD_STRMU	P49331 streptococc
18	65	61.9	1772	2 Q5SBN3_LACRE	Q5sbn3 lactobacill
19	64	61.0	1290	2 Q48756_LEUME	Q48756 leuconostoc
20	64	61.0	1506	2 Q56CX8_9STRE	Q56cx8 streptococc
21	64	61.0	1512	2 Q9WXJ5_9STRE	Q9wxj5 streptococc
22	63	60.0	1575	2 Q9LCH3_STROR	Q9lch3 streptococc
23	63	60.0	1577	2 Q54178_STRGN	Q54178 streptococc
24	63	60.0	1595	2 Q5SBN3_LACSK	Q5sbn3 lactobacill
25	63	60.0	1599	2 Q00599_STRSL	Q00599 streptococc
26	58	55.2	1449	2 Q68542_STRSL	Q68542 streptococc
27	58	55.2	1449	2 Q55264_STRSL	Q55264 streptococc
28	56	53.3	1577	2 Q55265_STRSL	Q55265 streptococc
29	55	52.4	382	2 Q8YWN3_ANASP	Q8ynw3 anabaena sp
30	54	51.4	1463	2 Q5SBN6_LACFE	Q5sbn6 lactobacill
31	53	50.5	1330	2 Q84CN4_LEUME	Q84cn4 leuconostoc

32 53 50.5 1477 2 Q9L466_LEUME Q9l466 leuconostoc
33 53 50.5 1508 2 Q9EZH5_LEUME Q9ezh5 leuconostoc
34 53 50.5 1508 2 Q52224_LEUME Q52224 leuconostoc
35 53 50.5 1781 2 Q5SBL9_LACRE Q5sbl9 lactobacill
36 53 50.5 1781 2 Q4JCS4_LACRE Q4jcs4 lactobacill
37 53 50.5 1781 2 Q4JLC7_LACRE Q4jlc7 lactobacill
38 51 48.6 659 2 Q8DR86_STRR6 Q8dr86 streptococc
39 51 48.6 798 2 Q9RKFO_STRCO Q9rkf0 streptomyc
40 51 48.6 1319 2 Q5JEU0_PYRKO Q5jeu0 pyrococcus
41 50 47.6 119 2 Q6UB60_9STRA Q6ub60 hyaloperono
42 50 47.6 1554 2 Q8KZL5_9STRE Q8kz15 streptococc
43 50 47.6 1772 2 Q5SBN0_LACRE Q5sbn0 lactobacill
44 49 46.7 119 2 Q7T4W3_9ENTO Q7t4w3 human echov
45 49 46.7 119 2 Q7T4W4_9ENTO Q7t4w4 human echov

ALIGNMENTS

RESULT 1

Q55263_9STRE PRELIMINARY; PRT; 1590 AA.
AC Q55263;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-JAN-1998 (TremBLrel. 05, Last sequence update)
DE 01-JUN-2003 (TremBLrel. 24, Last annotation update)
DE GTF-1.
GN Name-Glucosyltransferase;
OS Streptococcus sobrinus;
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OX Streptococcus.
OX NCBI_TaxID=1310;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 33478;
RA Sato S.;
RT "DNA and amino-acid sequences of water-insoluble-glucan synthetase produced from Streptococcus sobrinus ATCC 33478.";
RL Kagoeshima Daigaku Shigakubu Kiyo 16:23-29(1996).
RN [2]
RP PROTEIN SEQUENCE.
RX MEDLINE=91224988; PubMed=1827439;
RA Mooser G., Hefta S.A., Paxton R.J., Shively J.E., Lee T.D.;
RT "Isolation and sequence of an active-site peptide containing a catalytic aspartic acid from two Streptococcus sobrinus alpha-glucosyltransferases.";
RT J. Biol. Chem. 266:8916-8922(1991).
RL EMBL; D63570; BAA09792.1; -; Genomic_DNA.
DR PIR; A39841; A39841.
DR HSSP; P06653; 1GVM.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 3.
DR Pfam; PF02324; Glyco_hydro_70; 1.
SQ SEQUENCE 1590 AA; 176058 MW; 9DF7A3F2C6B4FD43 CRC64;

Query Match 100.0%; Score 105; DB 2; Length 1590;
Best Local Similarity 100.0%; Pred. No. 1e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVIANNVDKFSWGITDFEM 20
|||||
DB 847 VVIANNVDKFSWGITDFEM 866

RESULT 2

Q59983_9STRE PRELIMINARY; PRT; 1590 AA.
ID Q59983_9STRE
AC Q59983;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)

```
DE Glucosyltransferase-I precursor (EC 2.4.1.5).
GN Name=gtfI;
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1310;
RN [1]_TaxID=1310;
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OMZ176;
RX MEDLINE=91122498; PubMed=8312602;
RA Sato S., Inoue M., Handa N., Aizawa Y., Isobe Y., Katayama T.;
RT "DNA sequence of the glucosyltransferase gene of serotype d
RT Streptococcus sobrinus.";
RL DNA Seq. 4:19-27(1993).
RN [2]
RP PROTEIN SEQUENCE.
RX MEDLINE=91224988; PubMed=1827439;
RA Mooser G., Hefta S.A., Paxton R.J., Shively J.E., Lee T.D.;
RT "Isolation and sequence of an active-site peptide containing a
RT catalytic aspartic acid from two Streptococcus sobrinus alpha-
RT glucosyltransferases.";
RL J. Biol. Chem. 266:8916-8922(1991).
DR EMBL; D13858; BAA02976.1; -; Genomic_DNA.
DR PIR; A39841; A39841.
DR HSP; P06653; LHGX.
DR GO; GO:0047849; F:dextranucrase activity; IEA.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. .; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Glucosyltransferase; Signal; Transferase.
FT SIGNAL 1 38 Potential.
FT CHAIN 39 1590 Glucosyltransferase-I.
SQ SEQUENCE 1590 AA; 175956 MW; C3C83A57CF3C2B0E CRC64;

Query Match 100.0%; Score 105; DB 2; Length 1590;
Best Local Similarity 100.0%; Pred. No. 1e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVIANNVDFVSWGITDFEM 20
DB 847 VVIANNVDFVSWGITDFEM 866

RESULT 3
GTF2_STRDO
ID_GTF2_STRDO STANDARD; PRT; 1592 AA.
AC P27470;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
OS Streptococcus downei (Streptococcus sobrinus).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC STRAIN=6715 / Serotype G;
RX MEDLINE=91123227; PubMed=1704006;
RA Abo H., Matsumura T., Kodama T., Ohta H., Fukui K., Kato K.,
RA Kagawa H.;
RT "Peptide sequences for sucrose splitting and glucan binding within
RT Streptococcus sobrinus glucosyltransferase (water-insoluble glucan
RT synthetase).";
RL J. Bacteriol. 173:989-996(1991).
CC -!- FUNCTION: Production of extracellular glucans, that are thought to
CC play a key role in the development of the dental plaque because of
CC their ability to adhere to smooth surfaces and mediate the
CC aggregation of bacterial cells and food debris.
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-
```

```
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-
CC fructose + (1,6-alpha-D-glucosyl) (n+1).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes
CC water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
CC forms of glucans.
CC -!- SIMILARITY: Belongs to the glucosyl hydrolase 70 family.
CC -!- SIMILARITY: Contains 16 cell wall binding repeats.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC
CC EMBL; D90213; BAA14241.1; -; Genomic_DNA.
CC HSP; P06653; LGVM.
CC InterPro; IPR002479; Cell wall bd put.
CC InterPro; IPR003318; Glyco_hydro_70.
CC Pfam; PF01473; CW_binding_1; 3.
CC Pfam; PF02324; Glyco_hydro_70; 1.
KW Dental caries; Glucosyltransferase; Repeat; Signal; Transferase.
FT SIGNAL 1 38 Potential.
FT CHAIN 39 1592 Glucosyltransferase-I.
FT REPEAT 1093 1142 1.
FT REPEAT 1158 1207 2.
FT REPEAT 1222 1272 3.
FT REPEAT 1287 1337 4.
FT REPEAT 1402 1451 5.
FT REPEAT 1514 1563 6.
FT REPEAT 1577 1592 7 (incomplete).
FT REGION 39 1044 Catalytic (approximate).
FT REGION 1093 1592 7 X tandem repeats.
FT REGION 1093 1592 Glucan-binding (approximate).
SQ SEQUENCE 1592 AA; 176168 MW; BC0A66D079351ECF CRC64;

Query Match 100.0%; Score 105; DB 1; Length 1592;
Best Local Similarity 100.0%; Pred. No. 1e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVIANNVDFVSWGITDFEM 20
DB 847 VVIANNVDFVSWGITDFEM 866

RESULT 4
GTF1_STRDO
ID_GTF1_STRDO STANDARD; PRT; 1597 AA.
AC P11001;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
GN Name=gtfI;
OS Streptococcus downei (Streptococcus sobrinus).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1317;
RN [1]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC STRAIN=MFE28;
RX MEDLINE=87308014; PubMed=3040686;
RA Ferretti J.J., Gilpin M.L., Russell R.R.B.;
RT "Nucleotide sequence of a glucosyltransferase gene from Streptococcus
RT sobrinus MFE28.";
RL J. Bacteriol. 169:4271-4278(1987).
CC -!- FUNCTION: Production of extracellular glucans, that are thought to
CC play a key role in the development of the dental plaque because of
CC their ability to adhere to smooth surfaces and mediate the
CC aggregation of bacterial cells and food debris.
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-
```

fructose + (1,6-alpha-D-glucosyl) (n+1).
 -!- SUBCELLULAR LOCATION: Synthesized.
 -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both forms of glucans.
 -!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
 -!- SIMILARITY: Contains 19 cell wall binding repeats.

 This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.

EWBL; M17391;	RAC63063.1;	-;	Genomic DNA.
DR	InterPro;	IPR002479;	Cell_wall_bd_put.
DR	InterPro;	IPR003318;	Glyco_hydro_70.
DR	Pfam;	PF01473;	CW_binding_1;
DR	Pfam;	PF02324;	Glyco_hydro_70; 1.
KW	Dental caries;	Glucosyltransferase;	Repeat; Signal; Transferase.
FT	SIGNAL	38	Potential.
FT	CHAIN	36	Glucosyltransferase-I.
FT	REPEAT	1099	A repeat.
FT	REPEAT	1132	AC repeat.
FT	REPEAT	1163	AC repeat.
FT	REPEAT	1227	AC repeat.
FT	REPEAT	1292	AC repeat.
FT	REPEAT	1352	B repeat.
FT	REPEAT	1406	AC repeat.
FT	REPEAT	1465	AC repeat.
FT	REPEAT	1512	B repeat.
FT	REPEAT	1519	AC repeat.
FT	REPEAT	1582	AC repeat.
FT	REGION	39	A repeat (incomplete).
FT	REGION	1099	Catalytic (approximate).
FT	REGION	1597	1.25 A, 2 B and 5 AC repeats.
FT	REGION	1099	Bgp86A-binding (approximate).
SO	SEQUENCE	1597 AA:	Bgp86A-20086A798F_Crc64.
		177080 MW:	

Query Match	100.0%;	Score 105;	DB 1;	Length 1597;
Best Local Similarity	100.0%;	Pred. No. 1e-07;		
Matches 20;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	1	20
D6	853	872

```

RESULT 5
QBVOH3_STRMU QBVOH3_STRMU PRELIMINARY; PRT; 591 AA.
ID QBVOH3_STRMU
AC Q8VUH3;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Glucosyltransferase SI (Fragment).
DE Name=gtfc;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
[1]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=Xc;
RC MEDLINE=22322694; PubMed=12435673;
RX DOI=10.1128/AAC.46.12.3756-3764.2002;
RX Tsuda H., Yamashita Y., Shibata Y., Nakano Y., Koga T.;
RA "Genes involved in bacitracin resistance in Streptococcus mutans.";
RL Antimicrob. Agents Chemother. 46:3756-3764 (2002).
DR EMBL; AB078507; BAB83942.1; -; Genomic_DNA.
DR HSSP; P07762; 1M7X.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco hydro 70.

```

water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both forms of glucans.

-!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.

-!- SIMILARITY: Contains 5 cell wall binding repeats.

This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.

EMBL; M2054; AAA88592.1; -; Genomic DNA.
 DR EMBL; D8652; BAA26102.1; -; Genomic DNA.
 DR EMBL; D8655; BAA26106.1; -; Genomic DNA.
 DR EMBL; D8658; BAA26110.1; -; Genomic DNA.
 DR EMBL; D8661; BAA26114.1; -; Genomic DNA.
 DR EMBL; D8978; BAA26120.1; -; Genomic DNA.
 DR EMBL; AB014940; AAN58706.1; -; Genomic DNA.
 DR EMBL; M17361; AAA88589.1; -; Genomic DNA.
 DR PIR; JT0345; JT0345.
 DR HSP; P06653; IH8G.
 DR InterPro; IPR002479; Cell_wall_bd_put.
 DR InterPro; IPR003318; Glyco_hydro_70.
 DR Pfam; PF01473; CW_binding_1; 2.
 DR Pfam; PF02324; Glyco_hydro_70; 1.
 DR Complete proteome; Dental Caries; Glycosyltransferase; Repeat; Signal;
 KW Transferase.

1 34
 FT CHAIN 35 1455 Glucosyltransferase-SI.
 FT REPEAT 1126 1159 A repeat.
 FT REPEAT 1169 1200 A repeat.
 FT REPEAT 1227 1238 A repeat.
 FT REPEAT 1253 1303 AC repeat.
 FT REPEAT 1318 1330 A repeat (incomplete).
 FT REGION 35 1050 Catalytic (approximate).
 FT REGION 1126 1455 2.4 A, 1 C and 1 AC repeats.
 FT REGION 1126 1455 Glucan-binding (approximate).
 FT VARIANT 21 21 V -> I (in strain GS-5).
 FT VARIANT 81 81 P -> L (in strain MT4239).
 FT VARIANT 106 106 D -> V (in strain GS-5).
 FT VARIANT 116 116 S -> A (in strain GS-5 and strain MT4467).
 FT VARIANT 126 126 A -> T (in strain GS-5).
 FT VARIANT 150 151 SR -> PK (in strain GS-5, strain MT4239 and strain MT4467).
 FT VARIANT 256 256 A -> V (in strain GS-5 and strain MT4467).
 FT VARIANT 425 425 R -> N (in strain MT4251).
 FT VARIANT 519 519 Y -> D (in strain MT4245 and strain MT4251).
 FT VARIANT 538 538 R -> K (in strain MT4245 and strain MT4251).
 FT VARIANT 545 545 Y -> F (in strain MT4245 and strain MT4251).
 FT VARIANT 597 597 N -> D (in strain MT4245, strain MT4251, strain MT4467 and strain MT8148).
 FT VARIANT 600 600 R -> K (in strain MT4245, strain MT4251, strain MT4467 and strain MT8148).
 FT VARIANT 601 601 A -> T (in strain GS-5).
 FT VARIANT 614 614 M -> T (in strain GS-5).
 FT VARIANT 727 727 T -> I (in strain MT8148).
 FT VARIANT 734 734 A -> V (in strain MT8148).
 FT VARIANT 964 964 L -> F (in strain MT4239).
 FT VARIANT 1113 1113 N -> Y (in strain MT4239).
 FT VARIANT 1118 1118 I -> V (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).
 FT VARIANT 1204 1204 V -> I (in strain MT8148).
 FT VARIANT 1208 1208 DGH -> NGY (in strain GS-5, strain MT4467 and strain MT8148).
 FT VARIANT 1292 1294 Missing (in strain MT4245).
 FT VARIANT 1305 1369 I -> V (in strain MT8148).
 FT VARIANT 1326 1326 I -> V (in strain MT8148).
 FT VARIANT 1331 1331 T -> A (in strain GS-5, strain MT4239,

FT VARIANT 1377 1377 strain MT4467 and strain MT8148).
 FT VARIANT 1398 1398 R -> K (in strain MT8148).
 FT VARIANT 1424 1424 V -> I (in strain MT8148).
 FT VARIANT 1439 1439 D -> N (in strain MT4239).
 FT VARIANT 1439 1439 V -> I (in strain MT4239 and strain MT8148).
 FT VARIANT 1444 1444 S -> P (in strain MT8148).
 FT CONFLICT 1337 1455 QRLYFKSGVQAKGELITERKRIKYDPSNGEVNRVVR
 FT TSSGNNWYFGNDGYALIGHVVGRRVYFDENGVRVYASHD
 FT QRNWDYDRDFGRGSSAVFRHSRNGFDNFRF ->
 FT HASILSLMVFRLRESSLSQSVKVSNTMILIPMKFVIM
 FT (in Ref. 1).
 FT SQ SEQUENCE 1455 AA; 162966 MW; 3CB455A9A4FEC86 CRC64;
 Query Match 85.7%; Score 90; DB 1; Length 1455;
 Best Local Similarity 80.0%; Pred. No. 2.5e-05;
 Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVIANNVDKFSVSGITDFEM 20
 DB 876 VVIANNVDKFAEWGVTDFEM 895

RESULT 7
 ID GTFB_STRMU STANDARD; PRT; 1476 AA.
 AC P08987; O69381; O69384; O69387; O69390; O69396;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-MAY-2005 (Rel. 47, Last annotation update)
 DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
 DE (Sucrose 6-glucosyltransferase).
 GN Name=gtfI; OrderedLocusNames=SMU.1004;
 OS Streptococcus mutans.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1309;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=GS-5;
 RX MEDLINE=87308013; PubMed=3040685;
 RA Shiroza T., Ueda S., Kuramitsu H.K.;
 RT "Sequence analysis of the gtfB gene from Streptococcus mutans.";
 RL J. Bacteriol. 169:4263-4270(1987).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=MT4239 / Serotype C, MT4245 / Serotype e, MT4251 / Serotype f,
 RC MT4467 / Serotype e, and MT8148 / Serotype c;
 RX MEDLINE=98231643; PubMed=9570124;
 RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
 RA Kimura S., Hamada S.;
 RT "Molecular analyses of glucosyltransferase genes among strains of
 RT Streptococcus mutans.";
 RL FEMS Microbiol. Lett. 161:331-336(1998).
 RN [3]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RC STRAIN=UAI59 / ATCC 700610 / Serotype C;
 RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;
 RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
 RA Carson M.B., Prineaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
 RA Qian Y., Li S., Zhu H., Najjar P.Z., Lai H., White J., Roe B.A.,
 RA Ferretti J.J.;
 RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental
 RT pathogen.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
 CC -!- FUNCTION: Production of extracellular glucans, that are thought to
 CC play a key role in the development of the dental plaque because of
 CC their ability to adhere to smooth surfaces and mediate the
 CC aggregation of bacterial cells and food debris.
 CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-
 CC fructose + (1,6-alpha-D-glucosyl) (n+1).
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha

CC primer glucan unlike GTP-1.
 CC -!- MISCELLANEOUS: Synthesizes water-soluble glucans (alpha 1,6-glucose).
 CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
 CC -!- SIMILARITY: Contains 10 cell wall binding repeats.
 CC -----
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use as long as its content is in no way modified and this statement is not
 CC removed.
 CC -----
 CC EMBL; M30943; AAA26898.1; -; Genomic_DNA.
 DR HSSP; P06653; 1H8G.
 DR InterPro; IPR002479; Cell_wall_bd_put.
 DR InterPro; IPR003318; Glyco_hydro_70.
 DR Pfam; PF01473; CW_binding_1; 2.
 DR Pfam; PF02324; Glyco_hydro_70; 1.
 KW Dental caries; Glycosyltransferase; Repeat; Signal; Transferase.
 FT SIGNAL 1 36
 FT CHAIN 37 1365 Glucosyltransferase-S.
 FT REPEAT 157 177 Cell wall binding 1.
 FT REPEAT 178 197 Cell wall binding 2.
 FT REPEAT 1062 1082 Cell wall binding 3.
 FT REPEAT 1083 1102 Cell wall binding 4.
 FT REPEAT 1150 1169 Cell wall binding 5.
 FT REPEAT 1170 1190 Cell wall binding 6.
 FT REPEAT 1225 1243 Cell wall binding 7.
 FT REPEAT 1289 1308 Cell wall binding 8.
 FT REPEAT 1309 1328 Cell wall binding 9.
 FT REPEAT 1331 1352 Cell wall binding 10.
 FT REGION 198 1061 Catalytic (approximate).
 FT SEQUENCE 1365 AA; 151591 MW; 167296BSA2E8C476 CRC64;
 Query Match 65.7%; Score 69; DB 1; Length 1365;
 Best Local Similarity 73.7%; Pred. No. 0.061; Length 1365;
 Matches 14; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 2 VIANNVDFKFSWGITDFEM 20
 Db 837 VIANKGDLFKSWGITQFEM 855
 RESULT 13
 Q69A94 LEUME
 ID Q69A94 LEUME PRELIMINARY; PRT; 1454 AA.
 AC Q69A94
 DT 25-OCT-2004 (TrEMBLrel. 28, Created)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE Dextranase (EC 2.4.1.5).
 GN Name=derP;
 OS Leuconostoc mesenteroides.
 OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
 OX NCBI_TaxID=1245;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=IBT-PQ;
 RA Fernandez-Vazquez J.L., Lopez-Munguia A., Olvera C.;
 RT "Molecular characterization of a dextranase gene from Leuconostoc
 RT mesenteroides IBT-PQ isolated from pulque.";
 RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY504865; AAS79426.1; -; Genomic DNA.
 DR GO; GO:0047849; F:dextranase activity; IEA.
 DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
 DR GO; GO:0009250; P:glucan biosynthesis; IEA.
 DR InterPro; IPR002479; CW binding.
 DR InterPro; IPR003318; Glyco_hydro_70.
 DR Pfam; PF01473; CW_binding_1; 1.
 DR Pfam; PF02324; Glyco_hydro_70; 1.
 KW Glycosyltransferase; Transferase.
 FT SEQUENCE 1454 AA; 160569 MW; 2DC7342963B6F87 CRC64;

Query Match 65.7%; Score 69; DB 2; Length 1454;
 Best Local Similarity 68.4%; Pred. No. 0.066;
 Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 QY 1 VIANNVDFKFSWGITDFE 19
 Db 887 VQIANNTDLYKSWGITNFE 905
 RESULT 14
 Q6TXV4 LEUME
 ID Q6TXV4 LEUME PRELIMINARY; PRT; 1522 AA.
 AC Q6TXV4
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Dextranase (EC 2.4.1.5).
 GN Name=derX;
 OS Leuconostoc mesenteroides.
 OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
 OX NCBI_TaxID=1245;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=L0309;
 RA Wang J.-H., Shao Y.-C., Teng D., Yang Y.-L., Zhang F.;
 RT "Cloning and sequence analysis of gene coding for dextranase from
 RT Leuconostoc mesenteroides L0309.";
 RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AV743959; AAQ98615.2; -; Genomic DNA.
 DR GO; GO:0047849; F:dextranase activity; IEA.
 DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
 DR GO; GO:0009250; P:glucan biosynthesis; IEA.
 DR InterPro; IPR002479; CW binding.
 DR InterPro; IPR003318; Glyco_hydro_70.
 DR Pfam; PF01473; CW_binding_1; 2.
 DR Pfam; PF02324; Glyco_hydro_70; 1.
 KW Glycosyltransferase; Transferase.
 FT SEQUENCE 1522 AA; 169103 MW; 01BCC15468B913AB CRC64;
 Query Match 64.8%; Score 68; DB 2; Length 1522;
 Best Local Similarity 55.0%; Pred. No. 0.1;
 Matches 11; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
 QY 1 VIANNVDFKFSWGITDFEM 20
 Db 940 VVIAQNADQFKQGVTSFQL 959
 RESULT 15
 Q8KRE1 LEUME
 ID Q8KRE1 LEUME PRELIMINARY; PRT; 1527 AA.
 AC Q8KRE1
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Dextranase DsrD (EC 2.4.1.5).
 GN Name=derD;
 OS Leuconostoc mesenteroides.
 OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
 OX NCBI_TaxID=1245;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC MEDLINE=22573396; PubMed=12686639; DOI=10.1099/mic.0.26029-0;
 RA Neubauer H., Bauche A., Mollet B.;
 RT "Molecular characterization and expression analysis of the
 RT dextranase DsrD of Leuconostoc mesenteroides Lcc4 in homologous
 RT and heterologous Lactococcus lactis cultures.";
 RL Microbiology 149:973-982(2003).
 DR EMBL; AY017384; AAG61158.1; -; Genomic_DNA.
 DR HSSP; P06653; 1H8G.
 DR GO; GO:0047849; F:dextranase activity; IEA.
 DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
 DR GO; GO:0009250; P:glucan biosynthesis; IEA.

DR InterPro; IPR002479; CW_binding.
 DR InterPro; IPR003318; Glyco_Hydro_70.
 DR Pfam; PF01473; CW_binding_1; 2.
 DR Pfam; PF02324; Glyco_hydro_70; 1.
 KW Glycosyltransferase; Transferase.
 SQ SEQUENCE 1527 AA; 169835 MW; F9D0DE220BDB89668 CRC64;
 Query Match 64.8%; Score 68; DB 2; Length 1527;
 Best Local Similarity 55.0%; Pred. NO. 0.1;
 Matches 11; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
 QY 1 VVIANNVDFVSVSGITDEEM 20
 Db 945 VVIAQNADQFKQWGVTSFQL 964

Search completed: February 11, 2006, 19:39:14
 Job time : 2.69869 secs